

Mohamed Ben-Daya
Salih O. Duffuaa
Abdul Raouf
Jezdimir Knezevic
Daoud Ait-Kadi
Editors

Handbook of Maintenance Management and Engineering



Springer

Handbook of Maintenance Management and Engineering

Mohamed Ben-Daya • Salih O. Duffuaa
Abdul Raouf • Jezdimir Knezevic • Daoud Ait-Kadi
Editors

Handbook of Maintenance Management and Engineering

Mohamed Ben-Daya, Prof.
Salih O. Duffuaa, Prof.
King Fahd University of Petroleum
& Minerals (KFUPM)
Department of Systems Engineering
Dhahran 31261
Saudi Arabia
bendaya@ccse.kfupm.edu.sa
duffuaa@ccse.kfupm.edu.sa

Jezdimir Knezevic, Prof.
Mirce Academy
Longbrook Street
Exeter
Hems Mews
EX4 6AP
United Kingdom
mirce@mirce.net

Daoud Ait-Kadi, Prof.
Université Laval
Faculté de Sciences et de Génie
Département de Génie Mécanique
1314E Pavillon Adrien-Pouliot
Sainte-Foy, QC G1K 7P4
Canada
daoud.aitkadi@gmc.ulaval.ca

Abdul Raouf, Prof.Dr.
University of Management
and Technology
C-2 Johar Town
Lahore-54600
Pakistan
abdulraouf@umt.edu.pk

ISBN 978-1-84882-471-3

e-ISBN 978-1-84882-472-0

DOI 10.1007/978-1-84882-472-0

Springer Dordrecht Heidelberg London New York

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009931371

© Springer-Verlag London Limited 2009

Watchdog Agent[®] is a registered trademark of the Center for Intelligent Maintenance Systems, University of Cincinnati, PO Box 210072, Cincinnati, OH 45221, U.S.A. <http://www.imscenter.net>

D2B[™] is a trademark of the American College of Cardiology, Heart House, 2400 N Street, NW, Washington, DC 20037, U.S.A. <http://www.acc.org/>

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Cover design: eStudioCalamar, Figueres/Berlin

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

To be able to compete successfully both at national and international levels, production systems and equipment must perform at levels not even thinkable a decade ago. Requirements for increased product quality, reduced throughput time and enhanced operating effectiveness within a rapidly changing customer demand environment continue to demand a high maintenance performance.

In some cases, maintenance is required to increase operational effectiveness and revenues and customer satisfaction while reducing capital, operating and support costs. This may be the largest challenge facing production enterprises these days. For this, maintenance strategy is required to be aligned with the production logistics and also to keep updated with the current best practices.

Maintenance has become a multidisciplinary activity and one may come across situations in which maintenance is the responsibility of people whose training is not engineering. This handbook aims to assist at different levels of understanding whether the manager is an engineer, a production manager, an experienced maintenance practitioner or a beginner. Topics selected to be included in this handbook cover a wide range of issues in the area of maintenance management and engineering to cater for all those interested in maintenance whether practitioners or researchers.

This handbook is divided into 6 parts and contains 26 chapters covering a wide range of topics related to maintenance management and engineering.

Part I deals with maintenance organization and performance measurement and contains two chapters. Chapter 1 by Haroun and Duffuaa describes the maintenance organization objectives, the responsibilities of maintenance, and the determinants of a sound maintenance organization. In Chapter 2, Parida and Kumar address the issues of maintenance productivity and performance measurement. Topics covered include important performance measures and maintenance performance indicators (MPI), measurement of maintenance productivity performance and various factors and issues like MPI and MPM systems, MPI standard and MPIs use in different industries.

Part II contains an overview and introduction to various tools used in reliability and maintenance studies and projects. In Chapter 3, Ben-Daya presents basic statistical concepts including an introduction to probability and probability distributions, reliability and failure rate functions, and failure statistics. In Chapter

4, Ben-Daya provides an overview of several tools including failure mode and effect analysis, root cause analysis, the Pareto chart, and cause and effect diagram.

Part III contains three chapters related to maintenance control systems. Chapter 5 by Duffuaa and Haroun presents the essential elements and structure of maintenance control. Topics included cover required functions for effective control, the design of a sound work order system, the necessary tools for feedback and effective maintenance control, and the steps of implementing effective maintenance control systems. Cost control and budgeting is the topic of Chapter 6 by Mirghani. This chapter provides guidelines for budgeting and costing planned maintenance services. Topics covered include overview of budgeting and standard costing systems, budgeting framework for planned maintenance, a methodology for developing standard costs and capturing actual costs for planned maintenance jobs, and how detailed cost variances could be generated to assess the cost efficiency of planned maintenance jobs. The final chapter in this part is Chapter 7 by Riane, Roux, Basile, and Dehombreux. The authors discuss an integrated framework called OPTIMAIN that allows maintenance decision makers to design their production system, to model its functioning and to optimize the appropriate maintenance strategies.

Part IV focuses on maintenance planning and scheduling and contains five chapters. Forecasting and capacity planning issues are addressed in Chapter 8 by Al-Fares and Duffuaa. Topics covered include forecasting techniques, forecasting maintenance workload, and maintenance capacity planning. Necessary tools for these topics are presented as well and illustrated with examples. Chapter 9 by Diallo, Ait-Kadi and Chelbi deals with spare parts management. This chapter addresses the problem of spare parts identification and provisioning for multi-component systems. A framework considering available technical, economical and strategic information is presented along with appropriate mathematical models. Turnaround maintenance (TAM) is the object of Chapter 10 by Duffuaa and Ben-Daya. This chapter outlines a structured process for managing TAM projects. The chapter covers all the phases of TAM from its initiation several months before the event till the termination and writing of the final report. Chapter 11 by Al-Turki gives hands on knowledge on maintenance planning and scheduling for planners and schedulers at all levels. Topics covered include strategic planning in maintenance, maintenance scheduling techniques, and information system support available for maintenance planning and scheduling. Chapter 12 by Boukas deals with the control of production systems and presents models for production and maintenance planning. The production systems are supposed to be subject to random abrupt changes in their structures that may result from breakdowns or repairs.

Part V addresses maintenance strategies and contains eight chapters. Chapter 13 by Ait-Kadi and Chelbi presents inspection models. Topics covered include models for single and multi-component systems, and conditional maintenance models. Chapter 14 by Kothamasu, Huang and VerDuin offers a comprehensive review of System Health Monitoring and Prognostics. Topics surveyed include health monitoring paradigms, health monitoring tools and techniques, case studies, and organizations and standards. Ito and Nakagawa present applied maintenance models in Chapter 15. In this chapter, the authors consider optimal maintenance

models for four different systems: missiles, phased array radar, Full Authority Digital Electronic Control and co-generation systems based on their research. In Chapter 16, Siddiqui and Ben-Daya provide an introduction to reliability centered maintenance (RCM) including RCM philosophy, RCM methodology, and RCM implementation issues. Total productive maintenance (TPM) is the subject of Chapter 17 by Ahuja. Topics include basic elements of TPM, TPM methodology and implementation issues. Maintenance is an important concept in the context of warranties. Chapter 18 by Murthy and Jack highlights the link between the two subjects and discusses the important issues involved. Topics covered include link between warranty and maintenance, maintenance logistics for warranty servicing, and outsourcing of maintenance for warranty servicing. Delay Time (DT) Modeling for Optimized Inspection Intervals of Production Plant is the title of Chapter 19 by Wang. Topics covered include DT models for complex plant, DT model parameters estimation, and related developments and future research on DT modeling. Intelligent maintenance solutions and e-maintenance applications have drawn much attention lately both in academia and industry. The last chapter in Part V, Chapter 20 by Liyanage, Lee, Emmanouilidis and Ni deals with Integrated E-maintenance and Intelligent Maintenance Systems. Issues discussed include integrated e-maintenance solutions and current status, technical framework for e-maintenance, technology integration for advanced e-maintenance solutions, some industrial applications, and challenges of e-Maintenance application solutions.

Part VI deals with maintainability and system effectiveness and contains one chapter by Knezevic. It covers topics related to maintainability analysis and engineering and maintainability management.

Part VII contains five chapters presenting important issues related to safety, environment and human error in maintenance. Safety and maintenance issues are discussed in Chapter 22 by Pintelon and Muchiri. This chapter establishes a link between safety and maintenance, studies the effect of various maintenance policies and concepts on plant safety, looks at how safety performance can be measured or quantified, and discusses accident prevention in light of the safety legislation put in place by governments and some safety organizations. In Chapter 23, Raouf proposes an integrated approach for monitoring maintenance quality and environmental performance. Chapter 24 by Liyanage, Badurdeen and Ratnayake gives an overview of emerging sustainability issues and shows how the asset maintenance process plays an important role in sustainability compliance. It also elaborates on issues of quality and discusses best practices for guiding decisions. The last two chapters deal with human error in maintenance. Chapter 25 by Dhillon presents various important aspects of human reliability and error in maintenance. Finally Chapter 26 by Nicholas deals with human error in maintenance – a design perspective.

Maintenance professionals, students, practitioners, those aspiring to be maintenance managers, and persons concerned with quality, production and related areas will find this handbook very useful as it is relatively comprehensive when compared with those existing in the market.

The Editors

Acknowledgements

The editors would like to acknowledge the authors for their valuable contributions. This comprehensive handbook would not have been possible without their enthusiasm and cooperation throughout the stages of this project. We also would like to express our gratitude to all the reviewers who improved the quality of this book through their constructive comments and suggestions.

The editorial assistance of Atiq Siddiqui and Ali El-Rayyah with type setting and Blair Bremberg with English editing is highly appreciated. It takes a lot of patience to do all the typesetting and proofreading tasks necessary for such a project. Special thanks go to Dr. Sami Elferik for his enthusiastic support.

We are indebted to Simon Rees and Anthony Doyle of Springer and Sorina Moosdorf of le-tex publishing services oHG for their full cooperation and continuous assistance.

We would also like to express our gratitude to our families for their patience. Work on this book has sometimes been at the expense of their time.

Finally, we would like to acknowledge King Fahd University of Petroleum & Minerals for funding this project under the number SE/Maint.Mgt/331.

Contents

List of Contributorsxxv

Part I – Maintenance Organization

1 Maintenance Organization 3
Ahmed E. Haroun and Salih O. Duffuaa

- 1.1 Introduction.....3
- 1.2 Maintenance Organization Objectives and Responsibility5
- 1.3 Determinants of a Maintenance Organization.....6
 - 1.3.1 Maintenance Capacity Planning6
 - 1.3.2 Centralization vs Decentralization7
 - 1.3.2 In-house vs Outsourcing7
- 1.4 Design of the Maintenance Organization.....8
 - 1.4.1 Current Criteria for Organizational Change.....8
 - 1.4.2 Criteria to Assess Organizational Effectiveness9
- 1.5 Basic Types of Organizational Models9
- 1.6 Material and Spare Parts Management10
- 1.7 Establishment of Authority and Reporting13
- 1.8 Quality of Leadership and Supervision.....13
- 1.9 Incentives13
- 1.10 Education and Training.....14
- 1.11 Management and Labor Relations14
- 1.12 Summary15
- References.....15

2 Maintenance Productivity and Performance Measurement.....17
Aditya Parida and Uday Kumar

- 2.1 Introduction.....17
- 2.2 Performance Measurement and Maintenance Productivity.....19
- 2.3 Maintenance Performance21
- 2.4 Measurement of Maintenance Productivity23

2.4.1 Maintenance Performance Indicator (MPI).....	24
2.4.2 MPM Issues	24
2.4.3 MPM System	27
2.5 MPI Standards and MPIs as in Use in Different Industries	31
2.5.1 Nuclear Industry.....	32
2.5.2 Maintenance Indicators by EFNMS.....	33
2.5.3 SMRP Metrics.....	34
2.5.4 Oil and Gas Industry	35
2.5.5 Railway Industry.....	36
2.5.6 Process Industry	36
2.5.7 Utility Industry.....	37
2.5.8 Auto-industry Related MPIs for the CEO.....	38
2.6 Concluding Remarks.....	39
References.....	39

Part II – Methods and Tools in Maintenance

3 Failure Statistics45

Mohamed Ben-Daya

3.1 Introduction.....	45
3.2 Introduction to Probability.....	45
3.2.1 Sample Spaces and Events	45
3.2.2 Definition of Probability	46
3.2.3 Probability Rules.....	46
3.2.4 Conditional Probabilities.....	47
3.2.5 Random Variables.....	48
3.3 Probability Distributions.....	49
3.4 Reliability and Failure Rate Functions.....	51
3.4.1 Introduction.....	51
3.4.2 Reliability Function	52
3.4.3 Failure Rate Function.....	52
3.4.4 Mean Time Between Failure (MTBF)	53
3.5 Commonly Used Distributions.....	54
3.5.1 The Binomial Distribution	54
3.5.2 The Poisson Distribution.....	55
3.5.3 The Normal Distribution.....	56
3.5.4 The Lognormal Distribution	58
3.5.5 The Exponential Distribution.....	60
3.5.6 The Weibull Distribution	61
3.6 Failure Statistics.....	63
3.6.1 Types of Data.....	63
3.6.2 Parameter Estimation	64
References.....	73

4 Failure Mode and Effect Analysis75

Mohamed Ben-Daya

4.1 Introduction.....	75
-----------------------	----

4.2 FMEA Defined	76
4.3 FMEA Process	78
4.4 FMEA Applications	83
4.5 Related Tools	83
4.5.1 Root Cause Analysis	83
4.5.2 Pareto Chart	87
4.5.3 Cause and Effect Diagram	87
References	90

Part III – Maintenance Control Systems

5 Maintenance Control.....93

Salih O. Duffuaa and Ahmed E. Haroun

5.1 Introduction.....	93
5.2 The Maintenance Control Function	95
5.3 The Control Process.....	96
5.4 Functional Structure of Maintenance Control.....	97
5.5 Work Order System	99
5.5.1 Basic Documentation for Work Order System	99
5.5.2 Work Order System Flow	105
5.6 Tools Necessary for Effective Maintenance Control System	107
5.6.1 Work Control	108
5.6.2 Cost Control.....	109
5.6.3 Quality Control	109
5.6.4 Plant Condition Control	109
5.7 Effective Programs for Improving Maintenance Control.....	110
5.7.1 Emergency Maintenance.....	110
5.7.2 Reliability Improvement	110
5.7.3 Total Productive Maintenance	111
5.7.4 Computerized Maintenance Management and Information Technology	111
5.8 Summary.....	112
References.....	112

6 Guidelines for Budgeting and Costing Planned Maintenance Services115

Mohamed Ali Mirghani

6.1 Introduction.....	115
6.2 An Overview of Budgeting and Costing Systems.....	116
6.2.1 Budgeting Systems.....	116
6.2.2 Costing Systems.....	117
6.3 Proposed Budgetary System	118
6.3.1 Planned Maintenance Operating Budget.....	118
6.3.2 Financial Budget	119
6.3.3 The Budget Cycle	120
6.3.4 Top Management Support.....	120
6.3.5 Budget Performance Reports	122
6.4 Planned Maintenance Job Costing	123

6.4.1 Standard Cost Elements of a Planned Maintenance Job	123
6.4.2 Actual Cost Elements of a Planned Maintenance Job	128
6.4.3 Total Cost of a Planned Maintenance Job	129
6.4.4 Planned Maintenance Job Cost Variances	129
6.4.5 Significant Cost Variances	130
6.5 Summary and Conclusions	131
References	132
7 Simulation Based Approaches for Maintenance Strategies Optimization..	133
<i>Fouad Riane, Olivier Roux, Olivier Basile, and Pierre Dehombreux</i>	
7.1 Introduction	133
7.2 Reliability Models Estimation	134
7.2.1 Regression and ML Methods	134
7.2.2 Uncertainty Affecting Reliability Model	136
7.3 Maintenance Performance	138
7.3.1 Availability Model	138
7.3.2 Costs Model	139
7.4 Simulation Based Maintenance Framework	141
7.4.1 Toward a Unified Framework	141
7.4.2 Maintenance Strategies	142
7.4.3 Uncertainty Affecting Maintenance Performances	146
7.5 A Case Study	148
7.6 Conclusion	152
References	152
Part IV – Maintenance Planning and Scheduling	
8 Maintenance Forecasting and Capacity Planning	157
<i>Hesham K. Alfares and Salih O. Duffuaa</i>	
8.1 Introduction	157
8.2 Forecasting Basics	158
8.3 Qualitative Forecasting Techniques	159
8.3.1 The Delphi Method	160
8.4 Quantitative Forecasting Techniques	161
8.4.1 Simple Moving Averages	161
8.4.2 Weighted Moving Average	162
8.4.3 Regression Analysis	163
8.4.4 Exponential Smoothing	165
8.4.5 Seasonal Forecasting	167
8.4.6 Box-Jenkins Time Series Models	171
8.5 Error Analysis	172
8.6 Forecasting Maintenance Workload	173
8.7 Maintenance Capacity Planning	175
8.8 Deterministic Approaches for Capacity Planning	176
8.8.1 Modified Transportation Tableau Method	176
8.8.2 Mathematical Programming Methods	179
8.9 Stochastic Techniques for Capacity Planning	182

8.9.1 Queuing Models.....	182
8.9.2 Stochastic Simulation.....	186
8.10 Summary.....	188
References.....	189
9 Integrated Spare Parts Management.....	191
<i>Claver Diallo, Daoud Aït-Kadi, and Anis Chelbi</i>	
9.1 Introduction.....	191
9.2 Spare Parts Identification and Classification	192
9.3 Determination of the Required Quantity of Spare Parts.....	193
9.3.1 Recommendations.....	193
9.3.2 Reliability and Availability Based Procedures.....	193
9.3.3 Forecasting Procedure.....	198
9.3.4 Simulation.....	200
9.4 Inventory Control Policies	201
9.4.1 Model with Known and Constant Demand and Lead-time (EOQ Model).....	202
9.4.2 Model with Constant Demand and Perishable Items	202
9.4.3 Model with Random Demand and Lead-time	203
9.5 Joint Maintenance and Provisioning Strategies	204
9.5.1 Joint Replacement and Ordering Policy for a Spare Unit (One Unit Provisioning).....	205
9.5.2 Joint Replacement and Multiple Spare Parts Ordering Policy (Batch Provisioning).....	207
9.6 Inventory and Maintenance Policies for Reconditioned Spare Parts	209
9.6.1 Age of Recovered Parts to be Used for Replacement Actions.....	209
9.6.2 Review of Inventory Control Policies with Random Returns	213
9.7 Collaborative Management of Spare Parts.....	213
9.7.1 Access to Documentation and Knowledge Bases	213
9.7.2 Lead-time Reduction.....	214
9.7.3 Virtually Centralized Spare Parts Stock (Inventory Pooling)	214
9.7.4 Joint Replenishment of Spare Parts.....	216
9.8 Conclusion	218
References.....	218
10 Turnaround Maintenance.....	223
<i>Salih Duffuaa and Mohamed Ben-Daya</i>	
10.1 Introduction.....	223
10.2 Turnaround Initiation	225
10.3 Work Scope.....	226
10.4 Long Lead Time Resources	227
10.5 Contractors.....	228
10.6 TAM Planning	228
10.7 TAM Organization.....	229
10.8 Site Logistics.....	230
10.9 TAM Budget.....	230
10.10 Quality and Safety Plans.....	231

10.10.1 Quality Plan	231
10.10.2 Safety Plan	231
10.11 TAM Communication Procedures	232
10.12 TAM Execution	233
10.13 TAM Closing and Final Report	233
10.14 Conclusion	234
References.....	235
11 Maintenance Planning and Scheduling	237
<i>Umar M. Al-Turki</i>	
11.1 Introduction.....	237
11.2 Strategic Planning in Maintenance.....	240
11.3 Medium Range Planning.....	244
11.4 Short Range Planning	247
11.5 Maintenance Scheduling.....	247
11.5.1 Elements of Sound Scheduling	249
11.5.2 Maintenance Job Priority System.....	249
11.6 Scheduling Techniques	250
11.6.1 Gantt Charts and Scheduling Theory	250
11.6.2 Project Scheduling	253
11.6.3 Critical Path Method	256
11.6.4 Program Evaluation Review Techniques (PERT).....	258
11.7 Scheduling Using Computers.....	260
11.8 Summary	260
References.....	261
12 Models for Production and Maintenance Planning in Stochastic Manufacturing Systems	263
<i>E.K. Boukas</i>	
12.1 Introduction.....	263
12.2 Problem Statement and Preliminary Results	264
12.3 Dynamic Programming Approach	274
12.4 Linear Programming Approach	280
12.5 Conclusion	295
References.....	296
Part V – Maintenance Strategies	
13 Inspection Strategies for Randomly Failing Systems	303
<i>Anis Chelbi and Daoud Ait-Kadi</i>	
13.1 Introduction.....	303
13.1.1 Notation	304
13.2 Basic Inspection Model.....	305
13.2.1 Problem Definition.....	305
13.2.2 Working Assumptions and Mathematical Model.....	306
13.3 Extensions of the Basic Model.....	308
13.3.1 Inspection Models for Single Component Systems	308

13.4 Inspection Models for Multi-component Systems	318
13.4.1 Failure Tree Method Based Strategies	318
13.4.2 Cases of Cold and Hot Stand-by Systems with Known and Partially Known Lifetime Distributions.....	319
13.4.3 Case of Systems with Components Failure Dependency	320
13.5 Conditional Maintenance Models	321
13.5.1 Conditional Maintenance Models for Single Component Systems..	321
13.5.2 Conditional Maintenance Models for Multi-Component Systems...	329
13.6 Conclusion	331
References.....	332
 14 System Health Monitoring and Prognostics – A Review of Current Paradigms and Practices.....	 337
<i>Ranganath Kothamasu, Samuel H. Huang, William H. VerDuin</i>	
14.1 Maintenance Strategies: Motivations for Health Monitoring.....	337
14.2 Health Monitoring Paradigms	340
14.3 Health Monitoring Tools and Techniques.....	343
14.3.1 Reliability-based Maintenance.....	343
14.3.2 Model-based Approach to FDI.....	344
14.3.3 Signal-based FDI	346
14.3.4 Statistical FDI/Maintenance.....	347
14.4 Case Studies in System Monitoring and Control	347
14.5 Organizations and Standards.....	352
14.6 Summary and Research Directions	356
References.....	357
 15 Applied Maintenance Models	 363
<i>K. Ito and T. Nakagawa</i>	
15.1 Introduction.....	363
15.2 Missile Maintenance	365
15.2.1 Expected Cost	366
15.2.2 Optimal Inspection Policies	367
15.2.3 Numerical Illustrations.....	370
15.3 Phased Array Radar Maintenance.....	373
15.3.1 Cyclic Maintenance	374
15.3.2 Delayed Maintenance.....	377
15.3.3 Numerical Illustrations.....	379
15.4 Self-diagnosis for FADEC	380
15.4.1 Double Module System.....	381
15.4.2 Triple Module System.....	384
15.4.3 N Module System.....	386
15.4.4 Numerical Illustrations.....	386
15.5 Co-generation System Maintenance	387
15.5.1 Model and Assumptions.....	388
15.5.2 Analysis	389
15.5.3 Optimal Policy	390
15.5.4 Numerical Illustration	391
References.....	392

16 Reliability Centered Maintenance397

Atiq Waliullah Siddiqui and Mohamed Ben-Daya

16.1 Introduction.....	397
16.2 RCM Philosophy.....	400
16.2.1 RCM Principles and Key Features.....	400
16.2.2 RCM Goals and Benefits.....	401
16.2.3 System, System Boundary, Interfaces and Interactions.....	401
16.3 Failure and its Nature.....	404
16.4 RCM Methodology.....	405
16.4.1 Selecting Systems Selection and Collecting Information.....	405
16.4.2 System Boundary Definition.....	407
16.4.3 System Description and Functional Block Diagram.....	407
16.4.4 System Functions and Functional Failure.....	410
16.4.5 Failure Mode and Effective Analysis (FEMA).....	410
16.4.6 Logic or Decision Tree Analysis (LTA).....	411
16.4.7 Task Selection.....	411
16.5 RCM Implementation.....	411
16.5.1 Organizational Factors.....	412
16.5.2 RCM Teams.....	413
16.5.3 Scheduling Consideration and Training.....	413
16.6 Conclusion.....	414
References.....	414

17 Total Productive Maintenance417

P.S. Ahuja

17.1 Introduction to TPM.....	417
17.2 Evolution Towards TPM.....	420
17.3 Need of TPM.....	422
17.4 Basic Elements of TPM.....	424
17.5 Roadmap for TPM Implementation.....	429
17.6 An Ideal TPM Methodology.....	435
17.6.1 Introduction Phase (Phase I).....	435
17.6.2 TPM Initiatives Implementation Phase (Phase II).....	445
17.6.3 Standardization Phase (Phase III).....	451
17.7 Barriers in TPM Implementation.....	453
17.8 Success Factors for Effective TPM Implementation.....	456
17.9 Summary.....	458
References.....	458

18 Warranty and Maintenance461

D.N.P. Murthy and N. Jack

18.1 Introduction.....	461
18.2 Maintenance Modelling.....	462
18.2.1 Reliability.....	462
18.2.2 Types of Maintenance.....	462
18.2.3 Failure Modelling.....	462

18.3 Warranties.....	465
18.3.1 Base Warranties	465
18.3.2 Classification of Base Warranties.....	466
18.3.3 Warranty Servicing Cost Analysis.....	466
18.3.4 Extended Warranties.....	467
18.4 Link Between Warranty and Maintenance.....	467
18.4.1 Taxonomy for Classification.....	467
18.4.2 Warranty Servicing Involving Only CM.....	468
18.4.3 Warranty Servicing Involving Both CM and PM	469
18.5 Maintenance Logistics for Warranty Servicing	470
18.5.1 Strategic Issues.....	471
18.5.2 Tactical and Operational Issues	472
18.6 Outsourcing of Maintenance for Warranty Servicing.....	474
18.6.1 Agency Theory.....	474
18.7 Conclusions and Topics for Future Research.....	476
References.....	476
 19 Delay Time Modeling for Optimized Inspection Intervals of Production Plant.....	 479
<i>Wenbin Wang</i>	
19.1 Introduction.....	479
19.2 The DT Concept and Modeling Characteristics.....	480
19.3 The DT Models for Complex Plant.....	483
19.3.1 The Down Time/Cost Model	483
19.3.2 Modeling $E[N_f((i-1)T, iT)]$ and $E[N_s(iT)]$ Under the Assumption of Perfect Inspections.....	484
19.3.3 Modeling $E[N_f((i-1)T, iT)]$ and $E[N_s(iT)]$ Under the Assumption of Imperfect Inspections	485
19.4 Delay Time Model Parameters Estimation	487
19.4.1 Introduction.....	487
19.4.2 Complex System – Parameter Estimation.....	488
19.5 A Case Example.....	493
19.6 Other Developments in DT Modeling and Future Research Directions	496
References.....	497
 20 Integrated E-maintenance and Intelligent Maintenance Systems.....	 499
<i>Jayantha P. Liyanage, Jay Lee, Christos Emmanouilidis, and Jun Ni</i>	
20.1 Introduction.....	499
20.2 Condition-based Maintenance Technology and the State of Development.....	501
20.3 Integrated E-maintenance Solutions and Current Status	503
20.4 Technical Framework for E-maintenance.....	507
20.5 Watchdog Agent-based Intelligent Maintenance Systems	511
20.5.1 R ² M-PHM Platform	511
20.5.2 System Architecture.....	512

20.5.3 Toolbox for Multi-sensor Performance Assessment and Prognostics.....	514
20.5.4 Maintenance Decision Support System	518
20.6 Technology Integration for Advanced E-maintenance	520
20.6.1 Generic ICT Interface	520
20.6.2 Generic Interface Requirements for Watchdog Agents.....	525
20.6.3 Systems-user Interface Needs	528
20.7 Some Industrial Applications.....	528
20.7.1 E-maintenance Solutions for Complex Industrial Assets.....	528
20.7.2 Watchdog Technology for Product Life-cycle Design and Management.....	532
20.7.3 Watchdog Technology to Trouble-shoot Bearing Degradation	533
20.8 Challenges of E-maintenance Application Solutions.....	536
20.9 Conclusion	538
References.....	539

Part VI – Maintainability and System Effectiveness

21 Maintainability and System Effectiveness.....	547
<i>J. Knezevic</i>	
21.1 Introduction.....	547
21.2 The Concept of Maintainability	550
21.2.1 Maintainability Impact on System Effectiveness.....	552
21.2.2 Maintainability Impact on Safety	555
21.2.3 Undesirable Maintainability Practices	558
21.2.4 Desirable Maintainability Practices	559
21.3 Maintainability Analysis.....	561
21.3.1 Measures of Maintainability	565
21.3.2 Maintenance Labour-hour Factors	568
21.3.3 Maintenance Frequency Factors	569
21.3.4 Maintenance Cost Factors.....	570
21.3.5 Related Maintenance Factors	570
21.4 Empirical Data and Maintainability Measures.....	571
21.4.1 Possible Approaches to Analysis of Existing Data	571
21.4.2 Parametric Approach to Maintainability Data	572
21.4.3 Distribution Approach to Maintainability Data	572
21.4.4 Distribution Approach.....	575
21.5 Maintainability Engineering Predictions.....	576
21.5.1 Introduction.....	577
21.5.2 Concept of the Maintainability Block Diagram	577
21.5.3 Derivation of the Expression for the Maintainability Function.....	580
21.5.4 Maintainability Characteristics for Different Design Options	585
21.6 Maintainability Engineering Management.....	592
21.6.1 Role of the Maintainability Engineering Management Function.....	593
21.6.2 MEMF Opportunities.....	594
21.6.3 MEMF Obstacles	594
21.6.4 Design Methods for Attaining Maintainability	596

21.6.5 Maintainability Engineering Management – Lessons Learned	603
21.7 Concluding Remarks.....	607
References.....	610

Part VII – Maintenance Safety, Environment and Human Error

22 Safety and Maintenance.....613

Liliane Pintelon and Peter N. Muchiri

22.1 Setting the Scene.....	613
22.2 Definitions	615
22.2.1 Maintenance.....	615
22.2.2 Safety	616
22.2.3 Hazard.....	616
22.2.4 Stimuli.....	616
22.2.5 Accident	616
22.3 The Maintenance Link to Safety.....	617
22.3.1 The Role of Maintenance	617
22.3.2 Safety During Maintenance	620
22.3.3 Maintenance for Safety	622
22.3.4 Human Errors in Maintenance	625
22.3.5 Accident Causation Theories vs Maintenance	626
22.4 Maintenance Policies and Concepts vs Safety	629
22.4.1 Definitions	630
22.4.2 Maintenance Actions.....	630
22.4.3 Maintenance Policies	631
22.4.4 Maintenance Concepts	633
22.5 Maintenance Safety and Accident Prevention	636
22.5.1 Methods of Accidents and Hazards Avoidance in Maintenance	637
22.5.2 Analytical Approach	637
22.5.3 The Engineering Approach	638
22.5.4 Safety Culture	641
22.5.5 Safety Legislations.....	642
22.6 Safety Measurement.....	643
References.....	646

23 Maintenance Quality and Environmental Performance Improvement: An Integrated Approach.....649

Abdul Raouf

23.1 Introduction.....	649
23.2 Maintenance Quality.....	650
23.2.1 Improving Maintenance Quality	650
23.2.2 Benchmarking and Quality	651
23.2.3 Maintenance Audit.....	655
23.2.4 Improving Maintenance Quality Based on Stakeholder Feedback	656
23.3 Lean Manufacturing – Maintenance Quality Relationship	656
23.3.1 Basic Environmental Measure	656
23.4 Integrated Approach.....	660

23.5 Conclusion	663
References.....	663
24 Industrial Asset Maintenance and Sustainability Performance: Economical, Environmental, and Societal Implications.....	665
<i>Jayantha P. Liyanage, Fazleena Badurdeen, R.M. Chandima Ratnayake</i>	
24.1 Introduction.....	665
24.2 Industrial Activities and Sustainability Trends	666
24.3 Sustainability Performance in Perspective.....	668
24.4 Sustainability Performance Framework: From Business to Asset	671
24.5 Defining Maintenance Custodianship Within an Asset's Sustainability Performance	676
24.6 Generic Maintenance Impact Management Process	683
24.7 Adapting an Effective Asset Maintenance Practice for Sustainability	686
24.8 Conclusion	689
References.....	689
25 Human Reliability and Error in Maintenance.....	695
<i>B.S. Dhillon</i>	
25.1 Introduction.....	695
25.2 Terms and Definitions.....	695
25.3 Human Reliability and Error in Maintenance-Related Facts, Figures, and Examples	696
25.4 Occupational Stressors, Human Performance Effectiveness, and Human Performance Reliability Function	697
25.5 Human Error Occurrence Ways, Consequences, and Classifications, and Maintenance Error in System Life Cycle	700
25.6 Reasons for the Occurrence of Human Error in Maintenance and Top Human Problems in Maintenance	701
25.7 Mathematical Models for Performing Maintenance Error Analysis in Engineering Systems.....	702
25.7.1 Model I.....	703
25.7.2 Model II	705
25.8 Useful Guidelines to Reduce the Occurrence of Human Error in Maintenance.....	707
References.....	709
26 Human Error in Maintenance – A Design Perspective	711
<i>Clive Nicholas</i>	
26.1 Introduction.....	711
26.2 Human Error in Aircraft Maintenance	712
26.3 Significance of Maintenance Error	713
26.4 Design Impact.....	717
26.5 Analysis Required for Design Solutions	718
26.5.1 Maintenance Tasks.....	721
26.5.2 Maintenance Errors.....	722
26.5.3 Causal Factors.....	724

26.6 Design Strategies and Principles..... 726

 26.6.1 Appreciate the Maintainer’s Perspective of the Aircraft..... 729

 26.6.2 Design for the Aircraft Maintenance Environment..... 729

 26.6.3 Protect the Aircraft and Protect the Maintainer..... 731

 26.6.4 Avoid Complexity of Maintenance Tasks..... 732

 26.6.5 Enable Adequate Maintenance Access 732

 26.6.6 Positively Standardise and Positively Differentiate 733

 26.6.7 Build Error Detection into the Maintenance Process 734

26.7 Conclusion 734

References..... 735

Index..... 737

List of Contributors

P.S. Ahuja,
Punjabi University, India

Daoud Aït-Kadi,
Université Laval, Québec, Canada

Hesham K. Alfares,
King Fahd University of Petroleum & Minerals, Saudi Arabia

Umar M. Al-Turki,
King Fahd University of Petroleum & Minerals, Saudi Arabia

Fazleena Badurdeen,
University of Kentucky, USA

Olivier Basile,
Faculté Polytechnique de Mon, Belgium

Mohamed Ben-Daya,
King Fahd University of Petroleum & Minerals, Saudi Arabia

El-Kebir Boukas,
École Polytechnique de Montréal, Canada

Anis Chelbi,
École Supérieure des Sciences et Techniques de Tunis

Pierre Dehombreux,
Faculté Polytechnique de Mon, Belgium

B.S. Dhillon,
University of Ottawa, Canada

Claver Diallo,
Dalhousie University, Canada

Salih O. Duffuaa,
King Fahd University of Petroleum & Minerals, Saudi Arabia

Christos Emmanouilidis,
CETI/ATHENA Research and Innovation Centre, Greece

Ahmed Haroun,
King Fahd University of Petroleum & Minerals, Saudi Arabia

Samuel H. Huang,
University of Cincinnati, Cincinnati, USA

Kodo Ito,
Mitsubishi Heavy Industries, Ltd., Japan

Nat Jack,
University of Abertay Dundee, UK

J. Knezevic,
MIRCE Academy, UK

Ranganath Kothamasu,
University of Cincinnati, Cincinnati, USA

Uday Kumar,
Luleå University of Technology, Sweden

Jay Lee,
University of Cincinnati, USA

Jayantha P. Liyanage,
University of Stavanger, Norway

Mohamed Ali Mirghani,
King Fahd University of Petroleum & Minerals, Saudi Arabia

Peter Muchiri,
Katholieke University Leuven, Belgium

D.N.P. Murthy,
The University of Queensland, Australia

Toshio Nakagawa,
Aichi Institute of Technology, Japan

Jun Ni,
University of Cincinnati, USA

Clive Nicholas,
Mirce Academy, UK

Aditya Parida,
Luleå University of Technology, Sweden

Liliane Pintelon,
Katholieke University Leuven, Belgium

Abdul Raouf,
University of Management and Technology, Pakistan

R.M. Chandima Ratnayake,
University of Stavanger, Norway

Fouad Riane,
Faculté Polytechnique de Mon, Belgium

Olivier Roux,
Faculté Polytechnique de Mon, Belgium

Atiq Waliullah Siddiqui,
King Fahd University of Petroleum & Minerals, Saudi Arabia

William H. VerDuin,
Vertech LLC, Chagrin Falls, USA

Wenbin, Wang,
University of Salford, UK

Part I

Maintenance Organization

Maintenance Organization

Ahmed E. Haroun and Salih O. Duffuaa

1.1 Introduction

Organizing is the process of arranging resources (people, materials, technology *etc.*) together to achieve the organization's strategies and goals. The way in which the various parts of an organization are formally arranged is referred to as the organization structure. It is a system involving the interaction of inputs and outputs. It is characterized by task assignments, workflow, reporting relationships, and communication channels that link together the work of diverse individuals and groups. Any structure must allocate tasks through a division of labor and facilitate the coordination of the performance results. Nevertheless, we have to admit that there is no one best structure that meets the needs of all circumstances. Organization structures should be viewed as dynamic entities that continuously evolve to respond to changes in technology, processes and environment, (Daft, 1989 and Schermerhorn, 2007).

Frederick W. Taylor introduced the concept of scientific management (time study and division of labor), while Frank and Lilian Gilbreth founded the concept of modern motion study techniques. The contributions of Taylor and the Gilbreths are considered as the basis for modern organization management. Until the middle of the twentieth century maintenance has been carried out in an unplanned reactive way and for a long time it has lagged behind other areas of industrial management in the application of formal techniques and/or information technology. With realization of the impact of poor maintenance on enterprises' profitability, many managers are revising the organization of maintenance and have developed new approaches that foster effective maintenance organization.

Maintenance cost can be a significant factor in an organization's profitability. In manufacturing, maintenance cost could consume 2–10% of the company's revenue and may reach up to 24% in the transport industry (Chelson, Payne and Reavill, 2005). So, contemporary management considers maintenance as an integral function in achieving productive operations and high-quality products, while maintaining satisfactory equipment and machines reliability as demanded by

the era of automation, flexible manufacturing systems (FMS), “lean manufacturing”, and “just-in-time” operations.

However, there is no universally accepted methodology for designing maintenance systems, *i.e.*, no fully structured approach leading to an optimal maintenance system (*i.e.*, organizational structure with a defined hierarchy of authority and span of control; defined maintenance procedures and policies, *etc.*). Identical product organizations, but different in technology advancement and production size, may apply different maintenance systems and the different systems may run successfully. So, maintenance systems are designed using experience and judgment supported by a number of formal decision tools and techniques. Nevertheless, two vital considerations should be considered: strategy that decides on which level within the plant to perform maintenance, and hence outlining a structure that will support the maintenance; planning that handles day-to-day decisions on what maintenance tasks to perform and providing the resources to undertake these tasks.

The maintenance organizing function can be viewed as one of the basic and integral parts of the maintenance management function (MMF). The MMF consists of planning, organizing, implementing and controlling maintenance activities. The management organizes, provides resources (personnel, capital, assets, material and hardware, *etc.*) and leads to performing tasks and accomplishing targets. Figure 1.1 shows the role organizing plays in the management process. Once the plans are created, the management’s task is to ensure that they are carried out in an effective and efficient manner. Having a clear mission, strategy, and objectives facilitated by a corporate culture, organizing starts the process of implementation by clarifying job and working relations (chain of command, span of control, delegation of authority, *etc.*).

In designing the maintenance organization there are important determinants that must be considered. The determinants include the capacity of maintenance, centralization vs decentralization and in-house maintenance vs outsourcing. A number of criteria can be used to design the maintenance organization. The criteria include clear roles and responsibilities, effective span of control, facilitation of good supervision and effective reporting, and minimization of costs.

Maintenance managers must have the capabilities to create a division of labor for maintenance tasks to be performed and then coordinate results to achieve a common purpose. Solving performance problems and capitalizing on opportunities could be attained through selection of the right persons, with the appropriate capabilities, supported by continuous training and good incentive schemes, in order to achieve organization success in terms of performance effectiveness and efficiency.

This chapter covers the organizational structure of maintenance activities. Section 1.2 describes the organization objectives and the responsibilities of maintenance, followed by the determinants of a maintenance organization in Section 1.3. Section 1.4 outlines the design of maintenance organization and Section 1.5 presents basic models for organization. The description of function of material and spare parts management is given in Section 1.6, and Section 1.7 outlines the process of establishing authority. The role of the quality of leadership and supervision is presented in Section 1.8 followed by the role of incentives in

Section 1.9. Sections 1.10 and 1.11 present education and training, and management and labor relations, respectively. A summary of the chapter is provided in Section 1.12.

1.2 Maintenance Organization Objectives and Responsibility

A maintenance organization and its position in the plant/whole organization is heavily impacted by the following elements or factors:

- Type of business, *e.g.*, whether it is high tech, labor intensive, production or service;
- Objectives: may include profit maximization, increasing market share and other social objectives;
- Size and structure of the organization;
- Culture of the organization; and
- Range of responsibility assigned to maintenance.

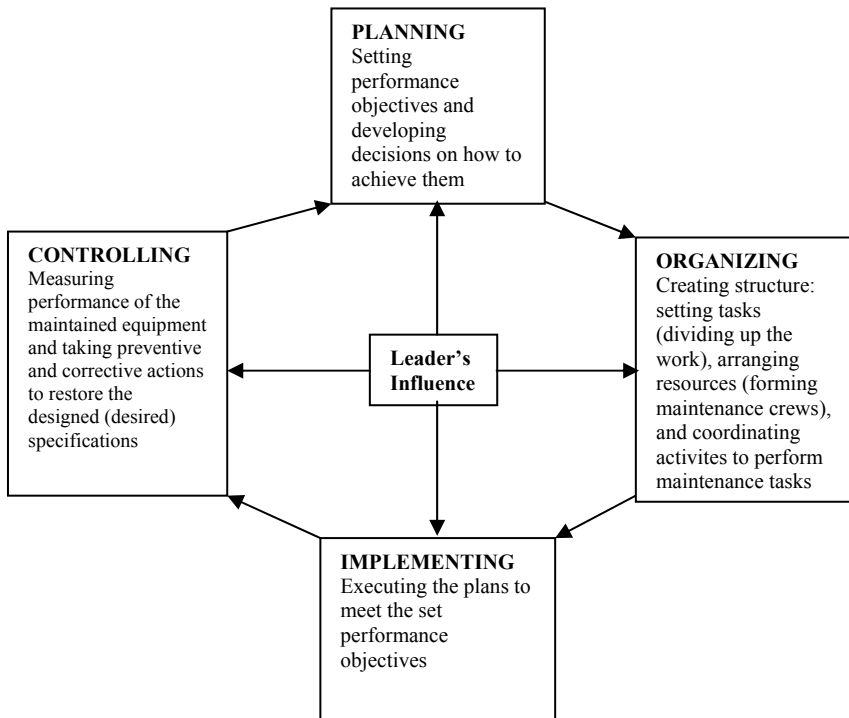


Figure 1.1. Maintenance organizing as a function of the management process

Organizations seek one or several of the following objectives: profit maximization, specific quality level of service or products, minimizing costs, safe and clean environment, or human resource development. It is clear that all of these

objectives are heavily impacted by maintenance and therefore the objectives of maintenance must be aligned with the objectives of the organization.

The principal responsibility of maintenance is to provide a service to enable an organization to achieve its objectives. The specific responsibilities vary from one organization to another; however they generally include the following according to Duffuaa *et al.* (1998):

1. Keeping assets and equipment in good condition, well configured and safe to perform their intended functions;
2. Perform all maintenance activities including preventive, predictive; corrective, overhauls, design modification and emergency maintenance in an efficient and effective manner;
3. Conserve and control the use of spare parts and material;
4. Commission new plants and plant expansions; and
5. Operate utilities and conserve energy.

The above responsibilities and objectives impact the organization structure for maintenance as will be shown in the coming sections.

1.3 Determinants of a Maintenance Organization

The maintenance organization's structure is determined after planning the maintenance capacity. The maintenance capacity is heavily influenced by the level of centralization or decentralization adopted. In this section the main issues that must be addressed when forming the maintenance organization's structure are presented. The issues are: capacity planning, centralization vs decentralization and in-house vs outsourcing.

1.3.1 Maintenance Capacity Planning

Maintenance capacity planning determines the required resources for maintenance including the required crafts, administration, equipment, tools and space to execute the maintenance load efficiently and meet the objectives of the maintenance department. Critical aspects of maintenance capacity are the numbers and skills of craftsmen required to execute the maintenance load. It is difficult to determine the exact number of various types of craftsmen, since the maintenance load is uncertain. Therefore accurate forecasts for the future maintenance work demand are essential for determining the maintenance capacity. In order to have better utilization of manpower, organizations tend to reduce the number of available craftsmen below their expected need. This is likely to result in a backlog of uncompleted maintenance work. This backlog can also be cleared when the maintenance load is less than the capacity. Making long run estimations is one of the areas in maintenance capacity planning that is both critical and not well developed in practice. Techniques for maintenance forecasting and capacity planning are presented in a separate chapter in this handbook.

1.3.2 Centralization vs Decentralization

The decision to organize maintenance in a centralized, decentralized or a hybrid form depends to a greater extent on the organization's philosophy, maintenance load, size of the plant and skills of craftsmen. The advantages of centralization are:

1. Provides more flexibility and improves utilization of resources such as highly skilled crafts and special equipment and therefore results in more efficiency;
2. Allows more efficient line supervision;
3. Allows more effective on-the-job training; and
4. Permits the purchasing of modern equipment.

However it has the following disadvantages:

1. Less utilization of crafts since more time is required for getting to and from jobs;
2. Supervision of crafts becomes more difficult and as such less maintenance control is achieved;
3. Less specialization on complex hardware is achieved since different persons work on the same hardware; and
4. More costs of transportation are incurred due to remoteness of some of the maintenance work.

In a decentralized maintenance organization, departments are assigned to specific areas or units. This tends to reduce the flexibility of the maintenance system as a whole. The range of skills available becomes reduced and manpower utilization is usually less efficient than in a centralized maintenance. In some cases a compromise solution that combines centralization and decentralization is better. This type of hybrid is called a cascade system. The cascade system organizes maintenance in areas and whatever exceeds the capacity of each area is challenged to a centralized unit. In this fashion the advantages of both systems may be reaped. For more on the advantages and disadvantages of centralization and decentralization see Duffuaa *et al.* (1998) and Niebel (1994).

1.3.3 In-house vs Outsourcing

At this level management considers the sources for building the maintenance capacity. The main sources or options available are in-house by direct hiring, outsourcing, or a combination of in-house and outsourcing. The criteria for selecting sources for building and maintaining maintenance capacity include strategic considerations, technological and economic factors. The following are criteria that can be employed to select among sources for maintenance capacity:

1. Availability and dependability of the source on a long term basis;
2. Capability of the source to achieve the objectives set for maintenance by the organization and its ability to carry out the maintenance tasks;

3. Short term and long term costs;
4. Organizational secrecy in some cases may be subjected to leakage;
5. Long term impact on maintenance personnel expertise; and
6. Special agreement by manufacturer or regulatory bodies that set certain specifications for maintenance and environmental emissions.

Examples of maintenance tasks which could be outsourced are:

1. Work for which the skill of specialists is required on a routine basis and which is readily available in the market on a competitive basis, *e.g.*,:
 - Installation and periodic inspection and repair of automatic fire sprinkler systems;
 - Inspection and repair of air conditioning systems;
 - Inspection and repair of heating systems; and
 - Inspection and repair of main frame computers *etc.*
2. When it is cheaper than recruiting your own staff and accessible at a short notice of time.

The issues and criteria presented in the above section may help organizations in designing or re-designing their maintenance organization.

1.4 Design of the Maintenance Organization

A maintenance organization is subjected to frequent changes due to uncertainty and desire for excellence in maintenance. Maintenance and plant managers are always swinging from supporters of centralized maintenance to decentralized ones, and back again. The result of this frequent change is the creation of responsibility channels and direction of the new organization's accomplishments *vs* the accomplishments of the former structure. So, the craftsmen have to adjust to the new roles. To establish a maintenance organization an objective method that caters for factors that influence the effectiveness of the organization is needed. Competencies and continuous improvement should be the driving considerations behind an organization's design and re-design.

1.4.1 Current Criteria for Organizational Change

Many organizations were re-designed to fix a perceived problem. This approach in many cases may raise more issues than solve the specific problem (Bradley, 2002). Among the reasons to change a specific maintenance organization's design are:

1. Dissatisfaction with maintenance performance by the organization or plant management;
2. A desire for increased accountability;

3. A desire to minimize manufacturing costs, so maintenance resources are moved to report to a production supervisor, thereby eliminating the (perceived) need for the maintenance supervisor;
4. Many plant managers are frustrated that maintenance seems slow paced, that is, every job requires excessive time to get done. Maintenance people fail to understand the business of manufacturing, and don't seem to be part of the team. This failure results in decentralization or distribution of maintenance resources between production units; and
5. Maintenance costs seem to rise remarkably, so more and more contractors are brought in for larger jobs that used to get done in-house.

1.4.2 Criteria to Assess Organizational Effectiveness

Rather than designing the organization to solve a specific problem, it is more important to establish a set of criteria to identify an effective organization. The following could be considered as the most important criteria:

1. Roles and responsibilities are clearly defined and assigned;
2. The organization puts maintenance in the right place in the organization;
3. Flow of information is both from top-down and bottom-up;
4. Span of control is effective and supported with well trained personal;
5. Maintenance work is effectively controlled;
6. Continuous improvement is built in the structure;
7. Maintenance costs are minimized; and
8. Motivation and organization culture.

1.5 Basic Types of Organizational Models

To provide consistently the capabilities listed above we have to consider three types of organizational designs.

- **Entralized maintenance.** All crafts and related maintenance functions report to a central maintenance manager as depicted in Figure 1.2. The strengths of this structure are: allows economies of scale; enables in-depth skill development; and enables departments (*i.e.*, a maintenance department) to accomplish their functional goals (not the overall organizational goals). This structure is best suited for small to medium-size organizations. The weaknesses of this structure are: it has slow response time to environmental changes; may cause delays in decision making and hence longer response time; leads to poor horizontal coordination among departments and involves a restricted view of organizational goals.
- **Decentralized maintenance.** All crafts and maintenance craft support staff report to operations or area maintenance as described in Figure 1.3. The strengths of this structure are that it allows the organization to achieve adaptability and coordination in production units and efficiency in a

centralized overhaul group and it facilitates effective coordination both within and between maintenance and other departments. The weaknesses of this structure are that it has potential for excessive administrative overheads and may lead to conflict between departments.

- Matrix structure, a form of a hybrid structure. Crafts are allocated in some proportion to production units or area maintenance and to a central maintenance function that supports the whole plant or organization as depicted in Figure 1.4. The strengths of this matrix structure are: it allows the organization to achieve coordination necessary to meet dual demands from the environment and flexible sharing of human resources. The weaknesses of this structure are: it causes maintenance employees to experience dual authority which can be frustrating and confusing; it is time consuming and requires frequent meetings and conflict resolution sessions. To remedy the weaknesses of this structure a management with good interpersonal skills and extensive training is required.

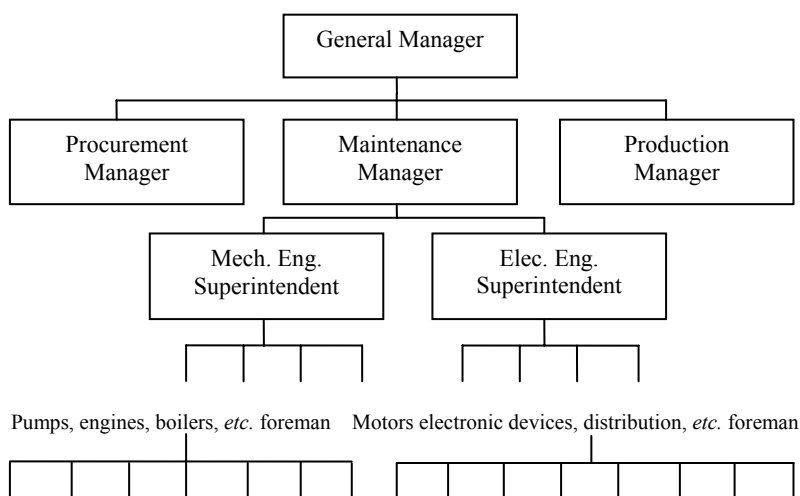
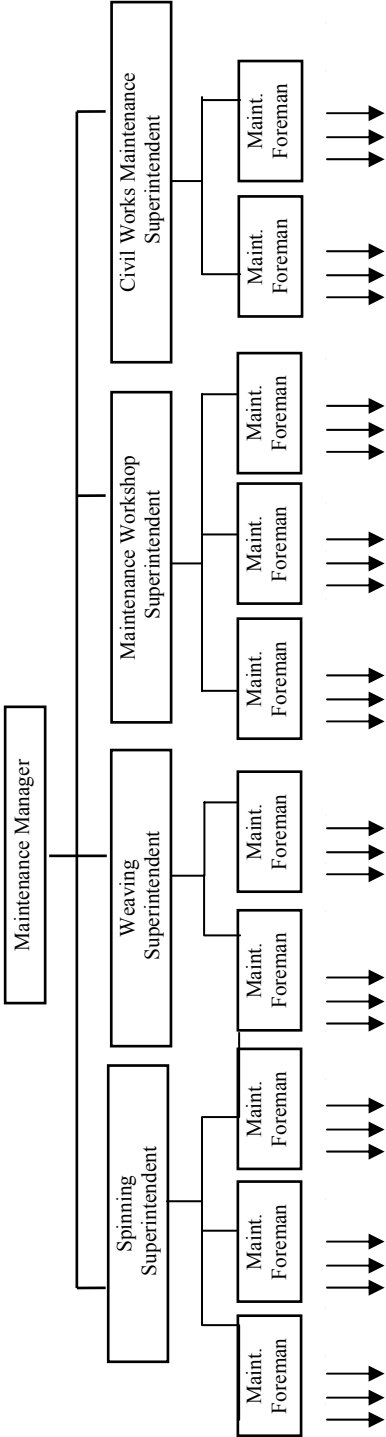


Figure 1.2. Centralized (functional) organizational structure

1.6 Material and Spare Parts Management

The responsibility of this unit is to ensure the availability of material and spare parts in the right quality and quantity at the right time at the minimum cost. In large or medium size organizations this unit may be independent of the maintenance organization; however in many circumstances it is part of maintenance. It is a service that supports the maintenance programs. Its effectiveness depends to a large extent on the standards maintained within the stores system. The duties of a material and spare parts unit include:



T E C H N I C I A N S

Figure 1.3. Functionally de-centralized organizational structure of maintenance in a textile factory

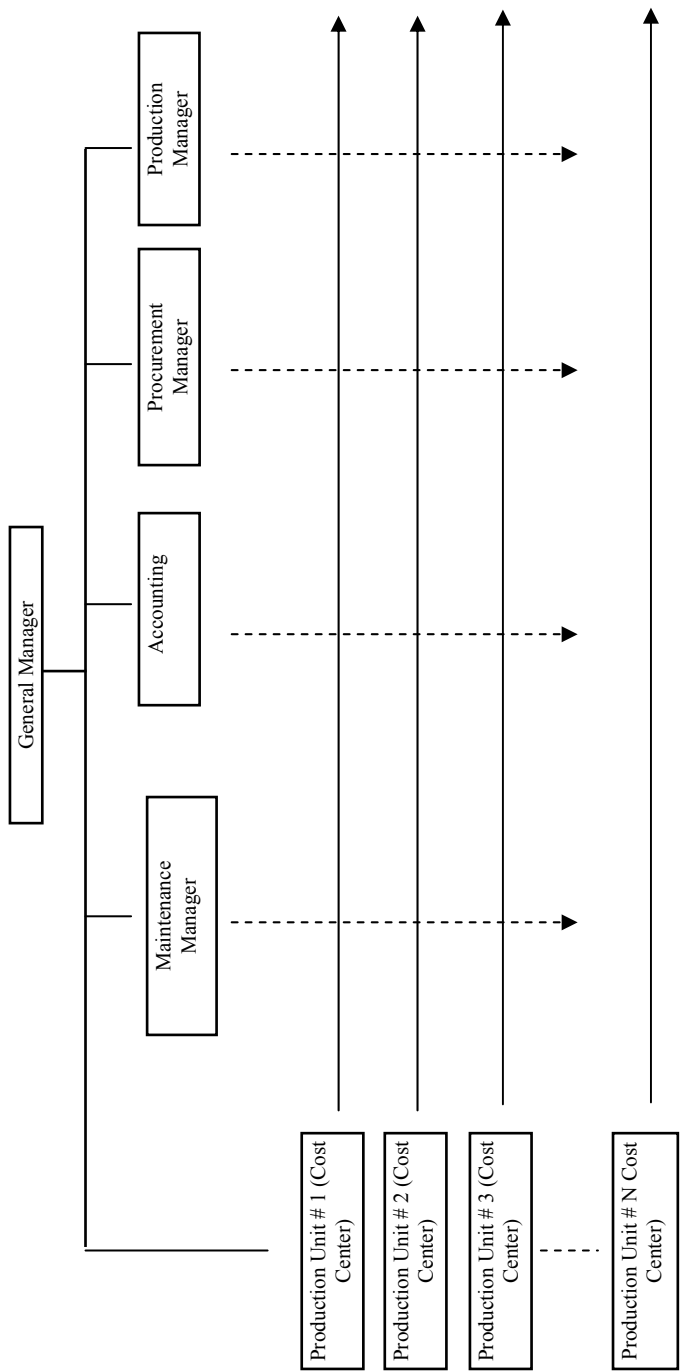


Figure 1.4. Matrix (de-centralized) organizational otruatures

1. Develop in coordination with maintenance effective stocking policies to minimize ordering, holding and shortages costs;
2. Coordinate effectively with suppliers to maximize organization benefits;
3. Keep good inward, receiving, and safe keeping of all supplies;
4. Issue materials and supplies;
5. Maintain and update records; and
6. Keep the stores orderly and clean.

1.7 Establishment of Authority and Reporting

Overall administrative control usually rests with the maintenance department, with its head reporting to top management. This responsibility may be delegated within the maintenance establishment. The relationships and responsibility of each maintenance division/section must be clearly specified together with the reporting channels. Each job title must have a job description prescribing the qualifications and the experience needed for the job, in addition to the reporting channels for the job.

1.8 Quality of Leadership and Supervision

The organization, procedures, and practices instituted to regulate the maintenance activities and demands in an industrial undertaking are not in themselves a guarantee of satisfactory results. The senior executive and his staff must influence the whole functional activity. Maintenance performance can never rise above the quality of its leadership and supervision. From good leadership stems the team-work which is the essence of success in any enterprise. Talent and ability must be recognized and fostered; good work must be noticed and commended; and carelessness must be exposed and addressed.

1.9 Incentives

The varied nature of the maintenance tasks, and differing needs and conditions arising, together with the influence of production activity, are not attuned to the adoption of incentive systems of payment. There are, however, some directions in which incentives applications can be usefully considered. One obvious case is that of repetitive work. The forward planning of maintenance work can sometimes lead to an incentive payment arrangement, based on the completion of known tasks in a given period, but care must be taken to ensure that the required standards of work are not compromised. In some case, maintenance incentives can be included in output bonus schemes, by arranging that continuity of production, and attainment of targets, provides rewards to both production and maintenance personnel.

1.10 Education and Training

Nowadays it is also recognized that the employers should not only select and place personnel, but should promote schemes and provide facilities for their further education and training, so as to increase individual proficiency, and provide recruits for the supervisory and senior grades. For senior staff, refresher courses comprise lectures on specific aspects of their work; they also encourage the interchange of ideas and discussion.

The further education of technical grades, craft workers, and apprentices is usually achieved through joint schemes, sponsored by employers in conjunction with the local education authority. Employees should be encouraged to take advantage of these schemes, to improve proficiency and promotion prospects.

A normal trade background is often inadequate to cope with the continuing developments in technology. The increasing complexity and importance of maintenance engineering warrants a marked increase in training of machine operators and maintenance craftsmen through formal school courses, reinforced by informed instruction by experienced supervisors.

The organization must have a well defined training program for each employee. The following provides guidelines for developing and assessing the effectiveness of the training program:

- Evaluate current personnel performance;
- Assess training need analysis;
- Design the training program;
- Implement the program; and
- Evaluate the program effectiveness.

The evaluation is done either through a certification program or by assessing the ability to achieve desired performance by persons who have taken a particular training program.

The implementation of the above five steps provides the organization with a framework to motivate personnel and improve performance.

1.11 Management and Labor Relations

The success of an undertaking depends significantly on the care taken to form a community of well-informed, keen, and lively people working harmoniously together. Participation creates satisfaction and the necessary team spirit. In modern industry, quality of work life (QWL) programs have been applied with considerable success, in the form of management conferences, work councils, quality circles, and joint conferences identified with the activities. The joint activities help the organization more fully achieve its purposes.

1.12 Summary

This chapter considered organizing as one of the four functions of management. It is the process of arranging resources (people, materials, technology, *etc.*) together to achieve the organization's strategies and goals. Maintenance organization structure is the way various part of the maintenance organization is formed including defining responsibilities and roles of units and individuals. A set of criteria are provided to assess and design organization structures and the main issues to be addressed are outlined. The issues include centralization, decentralization and outsourcing. The chapter describes three types of organization structures. In addition, several functions that could support maintenance organization such as material and spare management, training and the management of labor relations are presented.

References

- Bradley PS (2002) Designing the Best Maintenance. <http://www.samicorp.com>:1–5
- Chelson VJ, Payne CA, Reavill RP (2005) Management for Engineers Scientists and Technologists, 2nd edn. Wiley, Chichester England
- Daft LR (1989) Organization Theory and Design, 3rd edn. West Publishing Company, New York
- Duffuaa SO, Raouf A, Campbell JD (1998) Planning and Control of Maintenance Systems: Modeling and Analysis. Wiley, New York.
- Isa SEA (1989) Maintenance Management in Industrial Enterprises. J of Ind Coop 1:44–70.
- Haroun AE, Ogbugo MA (1981) Development of Maintenance Policies and Procedures for Crosfield Electronics at Milton Keynes. CIT, Cranfield, England
- Kanaway G (1992) Introduction to Work Study, 4th edn. ILO, Geneva
- Kelly A (1984) Maintenance Planning and Control, Butterworth, London
- Niebel B (1994) Engineering Maintenance Management, 2nd edn. Marcel Dekker, New York
- Schermerhorn JR (2007) Management, 9th edn. John Wiley, New York

Maintenance Productivity and Performance Measurement

Aditya Parida and Uday Kumar

2.1 Introduction

Maintenance productivity is one of the most important issues which govern the economics of production activities. However, productivity is often relegated to second rank, and ignored or neglected by those who influence production processes (Singh *et al.* 2000). Productivity in a narrow sense has been measured for several years (Andersen and Fagerhaug, 2007). Since maintenance activities are multi-disciplinary in nature with a large number of inputs and outputs, the performance of maintenance productivity needs to be measured and considered holistically with an integrated approach. With increasing awareness that maintenance creates added value to the business process; organizations are treating maintenance as an integral part of their business (Liyanage and Kumar, 2003). For many asset-intensive industries, the maintenance costs are a significant portion of the operational cost. Maintenance expenditure accounts for 20–50 % of the production cost for the mining industry depending on the level of mechanization. In larger companies, reducing maintenance expenditure by \$1 million contributes as much to profits as increasing sales by \$3 million (Wireman, 2007). The amount spent on the maintenance budget for Europe is around 1500 billion euros per year (Altmannshopfer, 2006) and for Sweden 20 billion euros per year (Ahlmann, 2002). In open cut mining, the loss of revenue resulting from a typical dragline being out of action is US \$ 0.5–1.0 million per day, and the loss of revenue from a 747 Boeing plane being out of action is roughly US \$ 0.5 million per day (Murthy *et al.* 2002). Therefore, the importance of maintenance productivity is understood more and more by the management of the companies.

There are several examples when lack of necessary and correct maintenance activities have resulted in disasters and accidents with extensive losses, like; Bhopal, Piper Alpha, space shuttle Columbia, power outages in New York, UK and Italy, during 2003. From asset management and changes in legal environment, the asset managers are likely to be charged with “corporate killing” due to changes in the legal environment for the future actions or omissions of the maintenance efforts (Mather, 2005). BP refinery in US paid a US \$21m fine and spent US \$1b for

repairs for an explosion at Texas City refinery, killing 15 and injured about 500 persons, making it the deadliest refinery accident (Bream, 2006). Prevention of such an accident could have enhanced BP's image besides saving a billion US \$. The measurement of maintenance performance has essentially become an essential element of strategic thinking for service and manufacturing industry. Due to outsourcing, separation of asset owners and asset managers, and complex accountability for the asset management, the measurement of asset maintenance performance and its continuous control and evaluation is becoming critical. As a result of the dramatic change in the use of technology, there is a growing reliance on software and professionals from other functional areas, for making or managing decisions on asset management and maintenance. Therefore, the performance of the maintenance process is critical for the long term value creation and economic viability of many industries. It is important that the performance of the maintenance process be measured, so that it can be controlled and monitored for taking appropriate and corrective actions to minimize and mitigate risks in the area of safety, meet societal responsibilities and enhance the effectiveness and efficiency of the asset maintained. A measure commonly used by industries is the maintenance performance for measuring the maintenance productivity.

In general, productivity is defined as the ratio of the output to input of a production system. The output of the production system is the products or services delivered while the input consists of various resources like the labour, materials, tools, plant and equipment, and others, used for producing the products or services. With a given input if more outputs of products or services can be produced, then higher productivity efficiency is achieved. Efficiency is doing the things right or it is the measure of the relationship of outputs to inputs and is usually expressed as a ratio. These measures can be expressed in terms of actual expenditure of resources as compared to expected expenditure of resources. They can also be expressed as the expenditure of resources for a given output. Effectiveness is doing the right things and measures the output conformance to specified characteristics.

Productivity is a combined measure for effectiveness and efficiency, *i.e.*, a productive organization is both effective and efficient. Measurement of productivity needs to consider various inputs and outputs of the products or services produced to be adequate and appropriate. Improvement in maintenance productivity can be achieved through reduction in maintenance materials as well as reductions in projects, outages and overhaul savings (Wireman, 2007). Production and service systems are heavily affected by their respective maintenance productivity. Maintenance systems operate in parallel to production systems to keep them serviceable and safe to operate at minimum cost. One way to reduce the operation cost and production cost is to optimize utilization of maintenance resources (Duffuaa and Al-Sultan, 1997), which enhances maintenance productivity. In order to measure the effectiveness of any maintenance system, we need to measure its productivity and identify the areas where improvements can be made (Raouf and Ben-Daya, 1995). Therefore, measuring maintenance productivity performance is critical for any production and operational company in order to measure, monitor, control and take appropriate and timely decisions. Since the cost of maintenance for different industries is substantial as compared to the

operational cost, more and more organizations are focussed to measure the performance of maintenance productivity.

The content of the chapter is as follows. After an introduction in Section 2.1, Section 2.2 discusses the performance measurement and maintenance productivity. In Section 2.3, maintenance performance and some of the important measures are explained. Section 2.4 deals with measurement of maintenance productivity performance and various factors and issues like MPI and MPM system. In Section 2.5, MPI standards and MPIs as in use at different industries are given followed by concluding remarks at Section 2.6.

2.2 Performance Measurement and Maintenance Productivity

Management needs information of maintenance performance for planning and controlling the maintenance process. The information needs to focus on the effectiveness and efficiency of the maintenance process, its activities, organization, cooperation and coordination with other units of the organization. Performance measurement (PM) has caught the imagination and involvement of researchers and managers from the industry alike, since the 1990s. With fast changes taking place in business and industry, the PM concepts and frameworks of past are outdated today, as they need to be modified as *per* today's requirements. Some of the concepts used in defining maintenance metrics are unclear regarding what to measure, how to communicate maintenance performance across the organization, aligning maintenance performance with objectives and strategies (Murthy *et al.* 2002). This essentially requires cascading down the corporate objectives into measurable targets up to shop floor level, and aggregating the measured maintenance performance indicators such as availability, reliability, mean time between failures, *etc.*, from shop floor level to the strategic levels for taking management decisions (Tsang, 2002). Murthy *et al.* (2002) mention that maintenance management needs to be carried out in both strategic and operational contexts and the organizational structure is generally structured into three levels. There is a need to identify and analyse various issues related to maintenance performance and to develop a framework which can address the related issues and challenges of maintenance management, maintenance performance measurement, performance measures and indicators. Maintenance and related processes across strategic, tactical and operational levels of hierarchy for the organization are required to be considered in the PM system. The performance measurement needs to be viewed along three dimensions (Andersen and Fagerhaug, 2007): (1) effectiveness: satisfaction of customer needs, (2) efficiency – economic and optimal use of enterprise resources and (3) changeability – strategic awareness to handle changes. Based on these three dimensions, a number of performance measures are developed. One example of the recent performance measurement system is the ENAPS (European Network for Advanced Performance Studies), a system based on a number of performance measures.

A PM system is defined as the set of metrics used to quantify the efficiency and effectiveness of actions (Neely *et al.* 1995). PM provides a general information basis that can be exploited for decision making purposes, both for management and

employees. Performance measurement is examined from three different levels, (1) from the individual performance measures, (2) from the system's performance measurement and (3) relationship between the PM system and its environment. Neely *et al.* (1995) also mentioned three PM concepts which highlight; classifications of performance measures as per their financial and non-financial perspectives, positioning the performance measures from the strategic context, and support of the organizational infrastructure, like resource allocation, work structuring, information system amongst others.

Maintenance Performance Measurement (MPM) is defined as "the multi-disciplinary process of measuring and justifying the value created by maintenance investment, and taking care of the organization's stockholder's requirements viewed strategically from the overall business perspective" (Parida, 2006). The MPM concept adopts the PM system, which is used for strategic and day to day running of the organization, planning, control and implementing improvements including monitoring and changes. PM is a means to measure the implementing strategies and policies of the management of the organization, which is the characteristic of MPM. Key performance indicator (KPI) is to be defined for each element of a strategic plan, which can break down to the PI at the basic shop floor or functional level. MPM linked to performance trends can be utilized to identify business processes, areas, departments and so on, that needs to be improved to achieve the organizational goals. Each organization is required to monitor and evaluate the need for performance improvement of the system. Thus, MPM forms a solid foundation for deciding where improvements are most pertinent at any given time. MPM can be effectively utilized for the improvement and the process evaluation and MPM data can also be used as a marketing tool, by providing information, like; quality and delivery time. MPM is also used as a basis for bench marking, in comparison to other organizations.

MPM is a powerful tool for aligning the strategic intent within the hierarchical levels of the entire organization. Thus, it allows the visibility of the company's goals and objectives from the CEO or strategic level to the middle management at tactical level and throughout the organization. MPM needs to be balanced from both financial and non-financial measures. Thus, MPM framework can be used for different purposes:

- A strategic planning tool;
- A management reporting tool;
- An operational control and monitoring tool; and
- A change management support tool.

A Performance Indicator (PI) is used for the measurement of the performance of any system or process. A PI compares actual conditions with a specific set of reference conditions (requirements) by measuring the distances between the current environmental situation and the desired situation (target), so-called 'distance to target' assessment (EEA, 1999). PIs should highlight opportunities for improvement within companies, when properly utilized (Wireman, 1998). PIs can be classified as leading or lagging indicators. Leading indicators provide an indication or warning of the performance condition in advance and act like a performance drivers. Non-financial indicators are examples of this. The lagging

indicators are mostly financial indicators which indicate performance after the activities are completed and hence are also called outcome measures. The outcome measures describe the resources spent or activities performed. Traditionally, management stresses profit measurement, which is mostly outcome measure based. The inputs or the resources put into an operation are mostly performance drivers, which need to be well controlled and managed for performance improvement. A good organizational system will combine the outcome measures with performance drivers as they are interrelated in a chain of ends and means. Within an organization, delivery time for the logistic department is an outcome measure, whereas for a customer it can be a performance driver for customer loyalty enhancement.

2.3 Maintenance Performance

Maintenance productivity aims at minimizing the maintenance cost dealing with the measurement of overall maintenance results/performance and maximizing the overall maintenance performance. Some of the measures of maintenance performances are availability, mean time between failures (MTTF), failure/breakdown frequency, mean time to repair (MTTR) and production rate index. Maintenance productivity indicators measures the usage of resources, like; labor, materials, contractors, tools and equipment. These components also form various cost indicators, such as man power utilization and efficiency, material usage and work order. Control of maintenance productivity (MP) ensures that the budgeted levels of maintenance efforts are being sustained and that required plant output is achieved (Kelly, 1997). Maintenance productivity deals with both maintenance effectiveness and efficiency.

For the process industry, machine downtime in the shop floor is one of the main issues for maintenance productivity. Unlike operational activities, maintenance activities are mostly non-repetitive in nature. Therefore, all maintenance personnel and managers face new problems with each breakdown or downtime of the plant or system, which needs multi-skill levels to solve the conflicting multi-objectives issues. For process or manufacturing industry, the product availability is given in Figure 2.1.

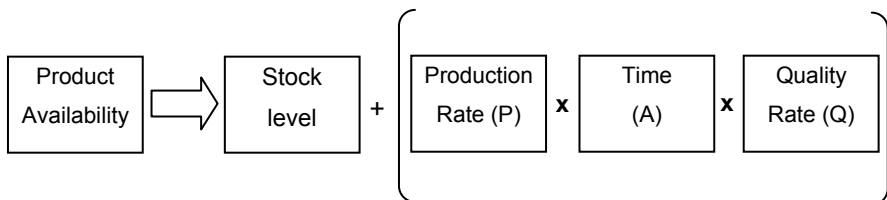


Figure 2.1. Elements of product availability (adapted with permission from Parida 2007)

For process or manufacturing industry, the input raw material issues are important as variation in quality of the raw material prevents the information of the quantity and quality of the products. This leads to reorder or recycle of the process to overcome the shortage of the required products, which also necessitates a safety

stock level. As given in Figure 2.1, the product availability is dependant on the production rate, available time for production and quality rate. The production rate is related to the plant or production capacity. If the maintenance effectiveness and efficiency is good, then the production rate will invariably be good. The availability time for production is also dependent on the repair or waiting time, *i.e.*, on the maintenance effectiveness. Quality of the product is also related to the number of stops, where quality loss is there during stop and start of the plant/system, besides the skill level of operators and the quality of the raw material *etc.* Thus, it can be seen that all the four parameters in product availability are dependent on maintenance directly or indirectly. The objective of the management of any process industry is to minimize the stock level, and increase the availability time, production and quality rate. The multiplication of the last three terms – availability, production and quality rate – provides the overall equipment effectiveness (OEE) figure which is one of the most important and effective key performance indicators (KPIs) in the performance measurement.

The machine breakdown or degradation of performance over time and accidents are some of the reason for the plant production interruption affecting the effectiveness of the plant. Normally, the production quantity is worked out by the management as *per* the market demand and situation. For achieving a greater market share, the management must be in a position to predict its plant capacity as well as improve it in a specified time.

The maintenance policy and safety performance of the plant plays a significant role in achieving the operational effectiveness of the plant. The management has to depend on the predicted plant capacity in order to meet the delivery schedules, cost, quality and quantity. An appropriate maintenance and safety strategy are required to be adapted for achieving the optimal production quantities.

Some of the important measures of maintenance productivity are:

- Total cost of maintenance/total production cost;
- A (availability) = (planned time - downtime)/planned time;
- P (production rate) = (standard time/unit)x(unit produced)/operating time; where; operating time = planned time – downtime;
- Q (quality rate) = (total production – defective quantity or number)/total production;
- Mean time to repair (MTTR) = sum of total repair time/number of breakdowns;
- Mean time between failure (MTBF) = number of operating hour/number of breakdowns;
- Maintenance breakdown severity = cost of breakdown repair/number of breakdown;
- Maintenance improvement = total maintenance manhours on preventive maintenance jobs ÷ total manhours available;
- Maintenance cost per hour = total maintenance cost/total maintenance man hours;
- Man power utilization = wrench time/total time;
- Manpower efficiency = time taken/planned time;

- Material usage/work order = total material cost/number of work order; and
- Maintenance cost index = total maintenance cost/total production cost.

All these measures of maintenance productivity need to be organization specific and defined accordingly. This is required to achieve a uniformity and transparency in understanding amongst all the employees and stakeholders of the organization, so that everyone speaks the same language. For example, for manpower utilization, wrench time needs to be specified for meaning and clarification.

2.4 Measurement of Maintenance Productivity

Various factors and issues are required to be considered for measurement of the maintenance productivity performance. Some of the important factors which need to be considered for measuring maintenance productivity are:

1. *The value created by the maintenance:* the most important factor in maintenance productivity measurement system is to measure the value created by maintenance process. As a manager, one must know that what is being done is what is needed by the business process, and if the maintenance output is not contributing/creating any value for the business, it needs to be restructured. This brings to the focus on doing the right things keeping in view the business objectives of the company.
2. *Revising allocations of resources:* the purpose for measuring the maintenance productivity effectiveness is to determine the additional investment requirement and to justify the investment made to the management. Alternatively, such measurement of activities also permits to determine the need for change of what is being done or how to do it more effectively by utilising the allocated resources.
3. *Health Safety and Environmental (HSE) Factors:* it is essential to understand the contribution of maintenance productivity towards HSE issues. An inefficient maintenance performance can lead to incidents and accidents (safety issue) and other health hazards, besides the environmental issues and encouraging an unhealthy work culture.
4. *Knowledge Management:* many companies focus on effective management of knowledge in their companies. Since technology is ever changing and is changing faster in the new millennium, this has brought in new sensors and embedded technology, information and communication technology (ICT) and condition based inspection technology like vibration, spectroscopy, thermography and others, which is replacing preventive maintenance with predictive maintenance. this necessitates a systematic approach for the knowledge growth in the specific field of specialization.
5. *New trends in operation and maintenance strategy:* companies need to adopt new operating and maintenance strategy in quick response to market demand, as well for the reduction of production loss and process waste. This strategy need to be continuously reviewed and modified.

6. Changes in Organizational Structure: organizations are trying to follow a flat and compact organizational structure, a virtual work organization, and empowered, self-managing, knowledge management work teams and work stations. Therefore a need exists to integrate the MPM system within the organization to provide a rewarding return for maintenance services.

2.4.1 Maintenance Performance Indicator (MPI)

Maintenance performance indicators (MPIs) are used for evaluating the effectiveness of maintenance carried out (Wireman, 1998). An indicator is a product of several metrics (measures). A performance indicator is a measure capable of generating a quantified value to indicate the level of performance, taking into account single or multiple aspects. The selection of MPIs depends on the way in which the MPM is developed. MPIs could be used for financial reports, for monitoring the performance of employees, customer satisfaction, the health, safety and environmental (HSE) rating, and overall equipment effectiveness (OEE), as well as many other applications. When developing MPIs, it is important to relate them to both the process inputs and the process outputs. If this is carried out properly, then MPIs can identify resource allocation and control, problem areas, the maintenance contribution, benchmarking, personnel performance, and the contribution to maintenance and overall business objectives (Kumar and Ellingsen, 2000).

2.4.2 MPM Issues

Each successful company measures their maintenance performance in order to remain competitive and cost effective in business. For improving maintenance productivity, it is essential that a structural audit is carried out, in which the following factors are evaluated (Raouf, 1994):

- Labor productivity;
- Organization staffing and policy;
- Management training;
- Planner training;
- Technical training;
- Motivation;
- Management control and budget;
- Work order planning and scheduling;
- Facilities;
- Stores, material and tool control;
- Preventive maintenance and equipment history;
- Engineering and condition monitoring;
- Work measurement and incentives; and
- Information system.

Understanding the need for MPM in the business and its work process, besides the associated issues, is critical for the development and successful implementation

of the maintenance productivity performance measurement. Besides maintenance process mapping, the associated issues are discussed.

Maintenance process mapping

It is essential to understand the maintenance process in detail before going on to study the issues involved in MPM system for any organization, so that implementation of the MPM system is possible without difficulty. The maintenance process starts with the maintenance objectives and strategy, which are derived from the corporate vision, goal and objectives based on the stakeholders' needs and expectations. Based on the maintenance objectives, maintenance policy, organization, resources and capabilities, a maintenance program is essentially developed. This program is broken down into different types of maintenance tasks. The execution of the maintenance tasks is undertaken at specified times and locations as *per* the maintenance plan. A maintenance task could be repair, replacement, adjustment, lubrication, modification or inspection. The management needs to understand the importance of maintenance and match the plan to the vision, goal and objectives of the organization. However, in real life there is a mismatch between the expectations of external and internal stakeholders and the capability between the organizational goals and the objectives and the resources allocated for maintenance planning, scheduling and between the execution and the reporting through data recording and analysis. There is a need to map the maintenance process and identify the gap between the maintenance planning and execution.

Logistic support, as *per* requirement is vital for maintenance planning, scheduling and execution. Such support includes the availability of spare parts, consumable materials, tools, instruction manuals, documents, *etc.* Logistic support acts as a performance driver which motivates and enhances the degree of maintenance performance. The non-availability of personnel, spares and consumable materials needs to be looked into, because otherwise it can act as a performance killer. Human factors such as unskilled and unwilling personnel act as a de-motivating factor which prevents the achievement of the desired results. Therefore, one must ensure the human resources and training necessary for the maintenance planning and execution team. The reporting system for MPM/MPIs is a major issue for any maintenance organization. It is necessary to understand the organizational need and then to procure or develop a system. The personnel using the MPM system need to be trained. Analysis of data plays an important role. It is equally important that the management should be involved in the whole process and there should be commitment and support from the top management.

The issues related to MPM are determined by answering the questions like:

- What indicators are relevant to the business and related to maintenance?
- How are the indicators related to one another and how do they take care of the stakeholders' requirements?
- Are the MPIs measurable objectively and how do the MPIs evaluate the efficiency and effectiveness of the organization?
- Are the MPIs challenging and yet attainable?
- Are the MPIs linked to the benchmarks or milestones quantitatively/qualitatively?

- How does one take decisions on the basis of the indicators?
- What are the corrective and preventive measures? and
- When and how does one update the MPIS?

The MPIS need to be developed based on the answers to the above questions. The relevant data need to be recorded and analyzed on a regular basis and used for monitoring, control of maintenance and related activities, and decision making for preventive and corrective actions. The MPIS could be time- and target-based, giving a positive or negative indication. An MPI could be trend-based in some cases. If it is positive or steady, meaning that everything is working well, and if it shows a negative trend and has crossed the lower limit of the target, then immediate decision to act urgently need to be taken. Various types of graphs and figures like a spider diagram could be used for indicating the health state of the technical system using different color codes for “excellent”, “satisfactory”, “improvement required” and “unsatisfactory performance level”. There could also be other visualization techniques using bar charts or other graphical tools for monitoring MPIS.

The issues related to the development and implementations of MPM are:

1. *Strategy*: how does one assess and respond to both internal and external stakeholders’ needs? How does one translate the corporate goal and strategy into targets and goals at the operational level, *i.e.*, converting a subjective vision into objective goals? How does one integrate the results and outcomes from the operational level to develop MPIS at the corporate level, *i.e.*, converting objective outcomes into strategic MPIS and linking them to strategic goals and targets? How to support innovation and training for the employees to facilitate an MPM-oriented culture?
2. *Organizational issues*: how to align the MPM system with the corporate strategy? Why there is a need to develop a reliable and meaningful MPM system? What should be measured, why it should be measured, how it should be measured, when it should be measured and what should be reported; when, how and to whom? How to establish accountability at various levels? How to improve communication within and outside the organization on issues related to information and decision making?
3. *How to measure?* how to select the right MPIS for measuring MPM? How to collect relevant data and analyze? How to use MPM reports for preventive and predictive decisions?
4. *Sustainability*: How to apply MPM strategy properly for improvement? How to develop an MPM culture across the organization? How to implement of a right internal and external communication system supporting MPM? How to review and modify the MPM strategy and system at regular intervals? How to develop and build trust in MPIS and MPM system at various levels?
5. *Specifying MPIS*: SMART test is frequently used to specify and determine the quality of the performance metrics (DOE-HDBK-1148- 2002). SMART stands for specific, measurable, attainable, realistic and timely.

The challenges associated with the development and implementation of an MPM system need to be considered for aligning it with the company’s vision and

goals. The performance measurement (PM) system needs to be aligned to organizational strategy (Kaplan and Norton, 2004; Eccles, 1991; Murthy *et al.* 2002). The balance scorecard of Kaplan and Norton (1992) focuses on financial aspects, customers, internal processes, and innovation and learning, for the first time, considering both the tangible and intangible aspects of the business. However, it did not consider the total effectiveness considering the external and internal effectiveness in a total, holistic and integrated manner. The total maintenance effectiveness is based on an organizational effectiveness model including both the external and the internal effectiveness. The concept of total maintenance effectiveness envelops the entire organization linking between the internal and external effectiveness. The total effectiveness is a product of the internal effectiveness characterised by issues related to effective and efficient use of resources to facilitate the delivery of the maintenance and related services to be delivered in the most effective way (engineering and business process related to planning and resource utilization) and external effectiveness characterised by customer satisfaction, growth in market share, *etc.* The performance measures for internal effectiveness is concerned with doing things in right way and can be measured in terms of cost effectiveness (maintenance costs per unit produced), productivity (number of work orders completed per unit time), *etc.* and deals with managing resources to produce services as *per* specifications.

The performance measures for external effectiveness deals with measures that have a long term effect on companies profitability and is characterised by doing right things, that is delivering services in a way (quality and timeliness) that meets customer requirements. Here the concept of delivering involves not only the services required by customers but also helping them in their other business process related to their own services. Such an attitude often helps in market growth and capturing or creating new markets. Whenever a balanced maintenance measurement system is developed, all the related criteria and parameters associated with the system are required to be examined. In any organization, first the maintenance process needs to be studied in detail and external effectiveness factors like the stakeholders requirements (front end processes) need to be understood. Then, based on the internal resources and capabilities, supply chain management (back end processes), the maintenance objectives and strategies are formulated, matching and integrating with that of the corporate ones. An important objective of the measurement system should be to bridge the gap and establish the relationship between the internal measures (causes) and the external measures (effects) (Jonsson and Lesshammar, 1999).

2.4.3 MPM System

An MPM system can be divided into three phases: the design of the performance measures, the implementation of the performance measures, and the use of the performance measures to carry out analysis/reviewing (Pun and White, 1996). The feedback from the reviewing to the system design keeps it valid in a dynamic environment.

Both the identification of appropriate measures and explicit consideration of trade-offs between them can be significantly assisted if the relationships between

measures are mapped and understood (Santos *et al.* 2002) well in advance. Therefore, the development of the MPM system requires the formation of a PM team which should include stakeholders at various levels and the management, and which should carry out preparatory work for this development work. The PM team should have clear and specified objectives, a time plan and a plan of action as pre-requisites.

2.4.3.1 Integration of Maintenance from Shop Floor to Strategic Level

The maintenance strategy should be derived from and integrated to the corporate strategy. In order to accomplish the top-level objectives of the espoused maintenance strategy, these objectives need to be cascaded into team and individual objectives. The adoption of fair processes is the key to successful alignment of these goals. It helps to harness the energy and creativity of committed managers and employees to drive the desired organizational transformations (Tsang, 1998). For a process industry or production system, the hierarchy is composed of the factory, process unit and component levels. The hierarchy corresponds to the traditional organizational levels of the top, middle and shop floor levels. However, there are some organizations which may require more than three hierarchical levels to suit their complex organizational structure. The MPM system needs to be linked to the functional and hierarchical levels for the meaningful understanding and effective monitoring and control of managerial decisions (Parida *et al.* 2005). Defining the measures and the actual measurements for monitoring and control constitute an extremely complex task for large organizations. The complexity of MPM is further increased for multiple criteria objectives, as shown in Figure 2.2.

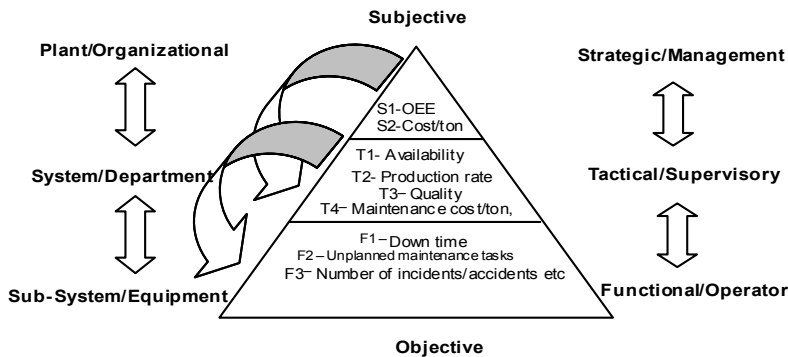


Figure 2.2. Hierarchical levels of an organization

From the hierarchical point of view, the top level considers corporate or strategic issues on the basis of soft or perceptual measures from stakeholders. In a way the strategic level is subjective, as it is linked to the vision and long-term goals (shown as S1 and S2 in Figure 2.3), though the subjectivity decreases down through the levels, with the highest objectivity existing at the functional level. The

second level considers tactical issues (shown as T1–T4 in Figure 2.3) such as financial and non-financial aspects both from the effectiveness and the efficiency point of view. This layer is represented by the senior or middle management, depending on the number of levels of the organization in question. If an organization has four hierarchical levels, then the second level represents the senior managerial level and the third level represents the managerial/supervisory level. The bottom level is represented by the operational personnel and includes the shop floor (shown as F1–F3 in Figure 2.2) engineers and operators. The corporate or business objective at the strategic level needs to be communicated down through the levels of the organization in such a way that this objective is translated into the language and meaning appropriate for the tactical or functional level of the hierarchy. The maintenance objectives and strategy, as derived from the stakeholders' requirements and corporate objectives and strategy, considering the total effectiveness, front-end processes and back-end processes, integrating the different hierarchical levels both from top-down and bottom-up manner involve the employees at all levels. At the functional level, the objectives are converted to specific measuring criteria. It is essential that all the employees speak the same language though out the entire organization.

2.4.3.2 Multi-criteria MPM System

The MPM system needs to facilitate and support the management leadership for timely and accurate decision making. The system should provide a solution for performance measurements linking directly with the organizational strategy and by considering both non-financial and financial indicators. At the same time, the system should be flexible, so as to change with time as and when required. The MPM system should be transparent and enable accountability for all the hierarchical levels. From the application and usage point of view, the MPM system should be technology user-friendly and should be facilitated by training the relevant personnel (Figure 2.3).

MPIs can be classified into seven categories (Parida *et al.* 2005) and are linked to each other for providing total maintenance effectiveness:

1. Customer satisfaction related indicators;
2. Cost related indicators;
3. Equipment related indicators;
4. Maintenance task related indicators;
5. Learning and growth related indicators;
6. Health safety and environment (HSE); and
7. Employee satisfaction related indicators.

Before implementation, the MPIs need to be tested for reliability, that is, the ability to provide the correct measures consistently over time, and for validity, which is the ability to measure what they are supposed to measure.

2.4.3.3 Implementation of the MPM System

Implementation of the developed MPM system for an organization is very critical. Neely *et al.* (2000) mention fear, politics and subversion as issues involved in this

phase. Ineffective use of information to improve operation without support of appropriate tools and lack of active management commitment and involvement is another critical issue, without which an MPM system cannot be effective or implemented fully (Santos *et al.* 2002). Dumond (1994) mentions lack of communication and dissemination of results as important issues for an MPM system. The alignment of PM with the strategic objectives of the organization at the design and development of MPM system is critical for achieving effectiveness of the implementation phase (Kaplan and Norton, 1992; Lynch and Cross, 1991).

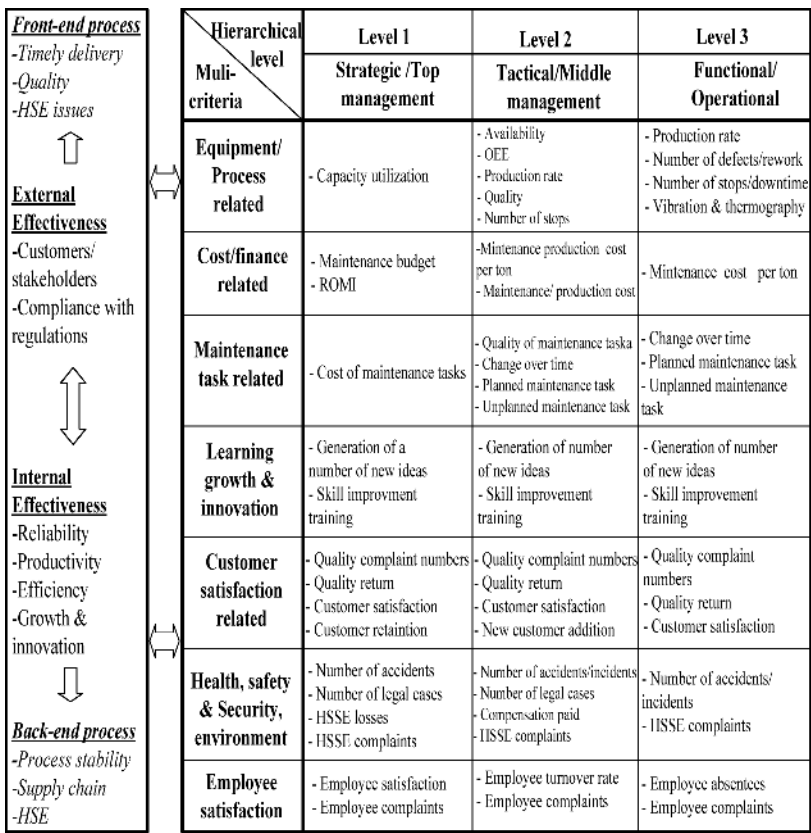


Figure 2.3. Multi-criteria frameworks for maintenance performance measurement (MPM) (Parida, 2006)

Prior to a pilot project studying the MPM system, it is desired that the relevant personnel of the organization should be trained in advance to create an awareness of MPM, the need for MPM and the benefits of MPM. A system of continuous monitoring, control and feedback needs to be institutionalized for the continuous improvement and successful implementation of the MPM system. A holistic view of a multi-criteria MPM framework showing the linkage of different MPis and criteria leading to achieving long term stakeholders' value is given in Figure 2.4.

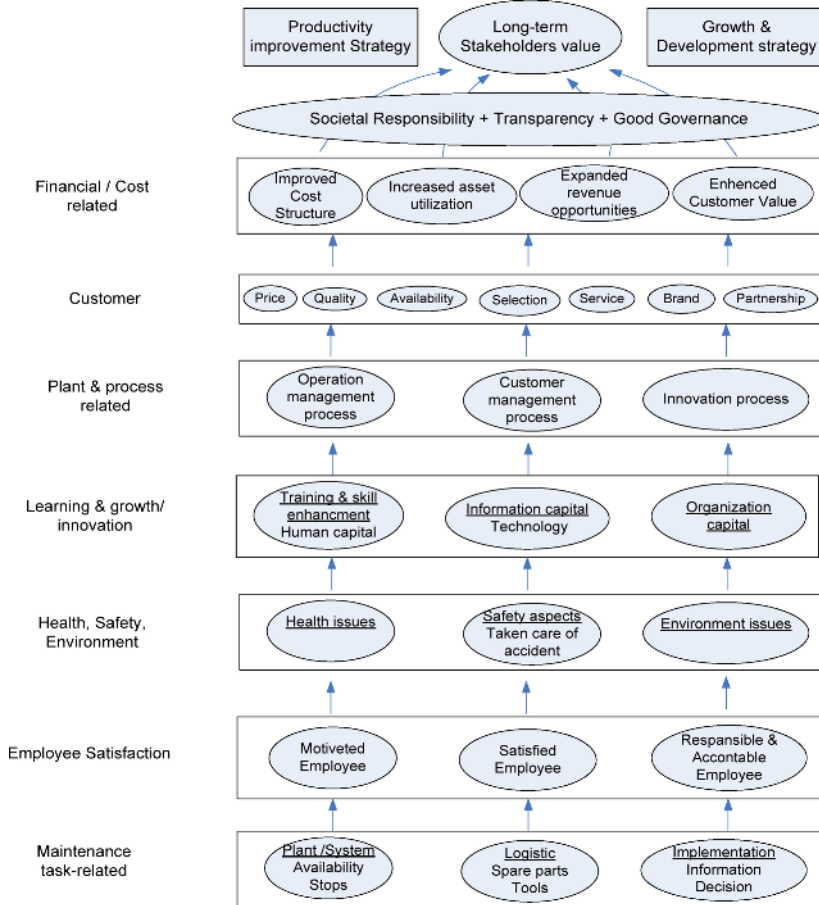


Figure 2.4. Holistic view of a multi-criteria MPM framework showing the linkage of different MPIs and criteria leading to achieving long term stakeholders' value (Parida (2006))

Thus, for implementing the MPM system, management and employee's commitments and involvement, communication and dissemination of results at each hierarchical level and MPIs' alignment with business objectives are some of the important issues need to be considered.

2.5 MPI Standards and MPIs as in Use in Different Industries

The greatest challenge for measuring maintenance performance is the implementation of the MPM system for validation of the MPIs under a real and industrial set up. Implementation first involves executing the plan and deploying the system developed in place of the previously existing or planned system. Second, it means operating with the selected measures and validating the assurance that the defined maintenance measurement system works on a day-to-day basis.

Without any formal measures of performance, it is difficult to plan control and improve the maintenance process. This is motivating senior business managers and asset owners to enhance the effectiveness of maintenance system. Also, with this, the focus is shifting to measure the performance of maintenance. Maintenance performance needs to be measured to evaluate, control and improve the maintenance activities for ensuring achievement of organizational goals and objectives. Different MPM frameworks and indicators to monitor, control and evaluate various performances are in use by different industries. More and more industries are working towards developing a specific MPM framework for their organization and identifying the indicators best suited to their industry. Organizations like International Atomic Energy Agency (IAEA) has already developed and published safety indicators during 2000 for nuclear power plants, and Society for Maintenance and Reliability Professionals (SMRP) and European Federation of National Maintenance Societies (EFNMS) have started organizing working groups and workshops to identify and select MPis for the industries. They have already defined and standardised some of the MPis to be followed by their associates and members. Besides, a number of industries have initiated research projects in collaboration with universities to identify suitable MPis as applicable to their specific industry. MPis are measures of efficiency, effectiveness, quality, timeliness, safety, and productivity amongst others. Some of the industries where MPM framework has been tried out are in the nuclear, oil and gas (O & G), railway, process industry and energy sectors amongst others. A different approach is used for developing the MPM framework and indicators for different industries, as per the stakeholders' requirements. Each organization under a specified industry is unique and as such the MPis and the MPM framework is required to be modified or developed specifically to meet its unique organizational and operational needs. Some of the MPM approaches, frameworks and MPis, as in use or under development by different societies, organizations and industries are discussed as under.

2.5.1 Nuclear Industry

The importance of the nuclear industry for energy generation as an alternate source is growing worldwide. International agencies like the International Atomic Energy Agency (IAEA) has been actively involved and sponsoring the development work in the area of indicators to monitor nuclear power plant (NPP) operational safety performance, from early 1990. The safe operation of nuclear power plants is the accepted goal for the management of the nuclear industry. A high level of safety results from the integration of the good design, operational safety and human performance. In order to be effective, a holistic and integrative approach is required to be adopted for providing a performance measurement framework and identifying the with desired safety attributes for the operation of the nuclear plant. Specific indicator trends over a period of time can provide an early warning to the management for investigating the causes of the observed change and comparing with the set target figure. Each plant needs to determine the indicators best suited to their individual needs, depending on the designed performance and, cost and benefit of operation/maintenance. The NPP performance parameters includes both

the safety and economic performance indicators, with overriding safety aspects. To assess the operational safety of NPP, a set of tools like the plant safety aspect (PSA), regulating inspection, quality assurance and self assessment are used. Two categories of indicators of commonly applied are; risk based indicators and safety culture indicators.

Operational Safety Performance Indicators

Indicator development starts attributes usage and the operational safety performance indicators are identified. Under each attribute, overall indicators are established for providing overall evaluation of relevant aspects of safety performance and, under each overall indicator, strategic indicators are identified. The strategic indicators are meant for bridging the gap between the overall and specific indicators. Finally, a set of specific indicators are identified/developed for each strategic indicators to cover all the relevant safety aspects of NPP. Specific indicators are used to measure the performance and identify the declining performance, so that management can take corrective decisions. Some of the indicators used in plants are given in Table 2.1 (IAEA 2000).

2.5.2 Maintenance Indicators by EFNMS

Since, 2004, European Federation of National Maintenance Societies (EFNMS) has conducted a number of workshops by forming a working group from amongst the member National Maintenance Societies of Europe resulting in identifying maintenance indicators for different industries for the national societies and branches. These workshops collected data for the maintenance indicators from industries and also trained the participants in the use of the indicators. The Croatian maintenance society (HDO) hosted the first workshop on maintenance indicators for the food and pharmaceutical business. The workshop was organised to train the maintenance managers in the use of maintenance indicators or Key Performance Indicators (KPIs) and to create an understanding of how to interpret the performance measured by the indicators. The participating maintenance managers were from the food and pharmaceutical industries. A number of workshops are organized in the same sector of industries to compare the results of the industry with the average maintenance performance in the sector. One of the important objectives of these workshops, besides the calculation of the indicators, is to increase the competence of the maintenance manager, who gets an understanding of the mechanism behind the indicators.

Table 2.1. Operational safety performance indicators

	Overall indicators	Strategic indicators	Specific indicators
1. Operates smoothly	1. Operating performance	1. Forced power reductions and outages	1. No. of forced power reductions and outages due to internal causes 2. No. of forced power reductions and outages due to external causes
	2. State of structures, systems and components	1. Corrective work orders issued	1. No. of corrective work orders issued for safety system 2. No. of corrective work orders issued for risk important BOP systems 3. Ratio of corrective work orders executed to work orders programmed 4. No. of pending work orders for more than 3 months
		2. Material condition	1. Chemistry Index (WANO performance indicators) 2. Ageing related indicators (condition indicators)
		3. State of the barriers	1. Fuel reliability (WANO) 2. RCS leakage 3. Containment leakage

These workshops resulted in the methodology for the use of the indicators and defined the draft EN standard 15341. The draft versions of the standard has 71 indicators to measure maintenance performance which are divided into economic indicators, technical indicators and organizational indicators. Among the indicators in the standard are the 13 indicators as defined by the working group of the European Federation of National Maintenance Societies in 2002. After approval, these indicators will be converted to EN standard. These activities resulted in developing a new European standard PrEN 15341 termed “Maintenance key performance indicators”, available at www.efnms.org/efnms/publications/Firstworkshopforfoodandpharmaceuticalbusiness.doc.

2.5.3 SMRP Metrics

The SMRP best practices committee has been chartered to identify and standardize maintenance and reliability metrics and terminology since 2004. They followed a six step process for the development of the metrics. The SMRP best practice metrics are published by the SMRP under the “Body of knowledge”, available for viewing at www.smrp.org. The numbering system for the metrics is explained on the web-page. Each metric has two files to describe the metric and feedback from the review of the metric. There are 45 metrics under development by different authors as of Feb 2006. A template is developed to provide a consistent method of describing each metric. The basic elements of each metric are:

Title	The name of the metric
Definition	A concise definition of the metric in easily understandable term
Objective	What the metric is designed to measure or report
Formula	A mathematical equation used to calculate the metric
Component definition	Clear definitions of each of the terms that are utilized in the metric formula
Qualifications	Guidance as to when or when not to apply the metric
Sample calculation	A sample calculation utilizing the formula with realistic values

A number of metrics are published at the SMRP web-site, which can be easily accessed. These metrics are explained in a clear and concise manner, which can be used by the personnel at different hierarchical level with out much difficulty. An example of the SMRP best practice metrics is given below:

2.5.4 Oil and Gas Industry

The cost of maintenance and its influence on the total system effectiveness of oil and gas industry is too high to ignore (Kumar and Ellingsen, 2000). The oil and gas industry uses MPIs and MPM framework extensively due to its ever growing and competitive nature of business, besides the productivity, safety and environmental issues. The safe operations of oil and gas production units are the accepted goal for the management of the industry. A high level of safety is essential from the integration of good design, operational safety and human performance. To be effective, an integrative approach is required to be adopted for providing an MPM framework and identifying the MPIs with desired safety attributes for the operation of the oil and gas production unit. Specific indicator trends over a period of time can provide an early warning to management to investigate the causes of the observed change and comparing with the set target figure. Each production unit needs to determine the indicators best suited to their individual needs, depending on the designed performance and cost and benefit of operation/maintenance. Some of the MPIs reported from plant level to result unit level to the result area for the Norwegian oil and gas industry are grouped into different categories as follows (Kumar and Ellingsen, 2000):

- *Production*
 - Produced volume oil (Sm3).
 - Planned oil-production (Sm3).
 - Produced volume gas (Sm3).
 - Planned gas-production (Sm3).
 - Produced volume condensate (Sm3).
 - Planned condensate- production (Sm3).

- *Technical integrity*
 - Backlog preventive maintenance (man-hours).
 - Backlog corrective maintenance (man-hours).
 - Bumber of corrective work orders.
- *Maintenance parameters*
 - Maintenance man-hours safety system.
 - Maintenance man-hours system.
 - Maintenance man-hours other systems.
 - Maintenance man-hours total.
- *Deferred production*
 - Due to maintenance (Sm3).
 - Due to operation (Sm3).
 - Due to drilling/well operations (Sm3).
 - Weather and other causes (Sm3).

2.5.5 Railway Industry

Railway operation and maintenance is meant for providing a satisfying service to the users, while meeting the regulating authorities' requirements. Today, one of the requirements for the infrastructure managers is to achieve cost effective maintenance activities, a punctual and cost-effective rail road transport system. As a result of a research project for the Swedish rail road transport system, the identified maintenance performance indicators are (Åhren and Kumar, 2004):

- Capacity utilization of infrastructure;
- Capacity restriction of infrastructure;
- Hours of train delays due to infrastructure;
- Number of delayed freight trains due to infrastructure;
- Number of disruptions due to infrastructure;
- Degree of track standard;
- Markdown in current standard;
- Maintenance cost per track-kilometer;
- Traffic volume;
- Number of accidents involving railway vehicles;
- Number of accidents at level crossings;
- Energy consumption per area;
- Use of environmental hazardous material;
- Use of non-renewable materials;
- Total number of functional disruptions; and
- Total number of urgent inspection remarks.

2.5.6 Process Industry

Measuring maintenance performance has drawn considerable interest in the utility, manufacturing and process industry in the last decade. Organizations are keen to know the return on investment made in maintenance spending, while meeting the

business objectives and strategy. Under challenges of increasingly technological changes, implementing an appropriate performance measurement system in an organization ensures that actions are aligned to strategies and objectives of the organization. Balanced, holistic and integrated multi-criteria hierarchical maintenance performance measurement (MPM) models developed with seven criteria and specific modification for the industry were tried out for implementation and achieving the total maintenance effectiveness for a pelletization plant and an energy producing service industry of Sweden (Parida *et al.* 2005). The MPIs for the process industry are:

1. Downtime (hours);
2. Change over time;
3. Planned maintenance tasks;
4. Unplanned tasks;
5. Number of new ideas generated;
6. Skill and improvement training;
7. Quality returned;
8. Employee complaints; and
9. Maintenance cost per ton.

In addition, MPIs identified for the multi-criterion hierarchical MPM framework, which are in existence and in use at LKAB (iron ore process company), are OEE, production cost per ton, planned maintenance tasks, quality complaints number, number of accidents, HSE complaints, and impact of quality.

2.5.7 Utility Industry

The MPIs for the utility industry in an energy sector will vary with that of other industries. The MPIs as identified for an energy sector organization of Europe are:

1. Customer satisfaction related: customer satisfaction is one of the main stakeholder group's requirements for the organization. Since, its customer is related to energy supply, duration and interruptions, and the contract, the customer satisfaction related MPIs are taken from the IEEE (1366-2003) and they are as under:

- SAIDI (system average interruption duration index), summation of customer interruption duration to total number of customer served;
- CAIDI (customer average interruption duration index, summation of customer interruption duration to total number of customer interrupted; and
- CSI (customer satisfaction index), obtained through customer survey.

2. Cost related: financial or cost is another main stakeholder group's requirements for any organization. Since, the total maintenance cost has to be controlled and the profit margin has to follow the Government's directive, these two MPIs are suggested to be included in the list of MPI:

- Total maintenance cost; and
- Profit margin.

3. *Plant/Process related*: the plant or process related MPIs also form important MPIs from internal stakeholder groups. Downtime of power generation and distribution, as well as the overall equipment effectiveness (OEE) rating of generation are the suggested MPIs from this group:

- Down time; and
- OEE rating (overall equipment effectiveness = availability × speed × quality).

4. *Maintenance task related*: the MPIs related to maintenance tasks are suggested as under:

- Number of unplanned stops (number and time);
- Number of emergency work; and
- Inventory cost.

5. *Learning and growth/innovation related*: the MPIs related to learning and growths, which are important for knowledge based organization, are:

- Number of new ideas generated; and
- Skill and improvement training.

6. *Health, safety and environment (HSE) related*: these are society related MPIs and very relevant to any organization today and they are:

- Number of accidents; and
- Number of HSE complaints.

7. *Employee satisfaction related*: employees are the most important internal stakeholders of the organization and their motivation, empowerment and accountability will be a supportive factor to achieve the organizational goal:

- Employee satisfaction level.

2.5.8 Auto-industry Related MPIs for the CEO

The MPIs used by an auto-industry are given in Table 2.2.

Table 2.2. The MPIs as used by an auto-industry for its CEO (Active strategy, 2006)

Financial	Increase profitability of core products	Core product profitability
	Increase sales of core models	Core model sales in m\$ Core model market share
Customer	Increase customer satisfaction	Customer satisfaction rating
Internal	Improve plant safety	Number of plant accidents
	Improve utilization of CRM system	% of CRM processes adopted
	Improve product launch effectiveness	% of launch plans on schedule
Learning and growth	Improve employee morale	Employee satisfaction survey
		Employee turnover

2.6 Concluding Remarks

Different MPM frameworks and indicators to monitor, control and evaluate maintenance productivity performance are in use by different industries. More and more industries are working towards developing specific MPM frameworks for their organization and identify the indicators best suited to their industry. Organizations like International Atomic Energy Agency (IAEA) have already developed and published safety indicators during 2000 for nuclear power plants, and Society for Maintenance and Reliability Professionals (SMRP) and European Federation of National Maintenance Societies (EFNMS) are organizing working groups and workshops to identify and select MPis for the industries. In addition, a number of industries have initiated research projects in collaboration with universities to identify suitable MPis as applicable to their specific industry. MPis are measures of efficiency, effectiveness, quality, timeliness, safety, and productivity amongst others. Some of the industries where MPM frameworks have been tried out are in the nuclear, oil and gas (O & G), railway, process industry and energy sector amongst others. A different approach is used for developing the MPM framework and indicators for different industries, as per the stakeholders' requirements. However, specific MPis are required to be identified and developed for an organization, which needs to be integrated with the MPM framework holistically.

References

- Active strategy (2006). Scorecard business review.
<http://www.activestrategy.com>.
- Ahlmann H (2002). From traditional practice to the new understanding: the significance of life cycle profit concept in the management of industrial enterprises, Proceedings of the International Foundation for Research in Maintenance, Maintenance Management & Modelling, 6–7 May, Växjö, Sweden.
- Åhren T and Kumar U (2004) Use of maintenance performance indicators: a case study at Banverket. Conference proceedings of the 5th Asia-Pacific Industrial Engineering and Management Systems Conference (APIEMS2004). Gold Coast, Australia
- Altmannshoffer R (2006) Industrielles FM, Der Facility Manager (In German), April Issue, pp. 12–13.
- Andersen B and Fagerhaug T (2002) Performance Measurement Explained, ASQ Quality Press, Milwaukee, Wisconsin.
- Andersen B and Fagerhaug T (2007) Performance measurement of Logistic processes, <http://www.prestasjonsledelse.net/publikasjoner/Performance%20measurement%20of%20logistics%20processes.pdf>, Accessed 16 April 2007.
- Bream R (2006) More fuel for anti-BP sentiment, Financial Times, London, DOE-HDBK-1148–2002.(2002)Work Smart Standard (WSS) Users' Handbook, Department of Energy, USA.
<http://www.eh.doe.gov/tecgststds/standard/hdbk1148/hdbk11482002.pdf>. Accessed June 30 2002

- Duffuaa SO and Al-Sultan KS (1997) Mathematical programming approaches for the management of the maintenance planning and scheduling. *J of Qual in Maint Eng* 3(3): 163–76.
- Dumond EJ (1994) Making best use of performance measures and information. *Int J of Oper and Prod Manag* 14(9):16–31.
- Eccles RG (1991) The performance measurement manifesto. *Harvard Business Review*, January-February, pp 131–137.
- EEA (European Environment Agency) (1999) Environmental indicators: Typology and overview. Technical Report No 25, Copenhagen.
- IAEA (2000) International Atomic Energy Agency. A Framework for the Establishment of Plant specific Operational Safety Performance Indicators, Report, Austria
- IEEE Standards 1366 (2003) IEEE Guide for Electric Power Distribution Reliability Indices, The Institute of Electrical and Electronics Engineers, Inc, New York, USA, 14 May, 2004.
- Jonsson P, Lesshammar M (1999) Evaluation and improvement of manufacturing performance measurement systems: the role of OEE. *Int J of Oper & Prod Manag* 19(1): 55–78.
- Kaplan RS and Norton DP (1992) The balanced scorecard – measures that drive performance. *Harvard Business Review*, pp 71–79.
- Kaplan R S and Norton DP (2004) Strategy Map: Converting intangible asset into tangible outcomes, Harvard Business School Press, USA
- Kelly A (1997) Maintenance Organization and System., Butterworth Heinemann, UK.
- Kumar U and Ellingsen HP (2000) Development and implementation of maintenance performance indicators for the Norwegian oil and gas industry. Proceedings of the 14th International Maintenance Congress (Euro maintenance 2000), 7–10 March 2000, Gothenburg, Sweden, pp 221–228.
- Liyanaage JP and Kumar U (2003) Towards a value-based view on operations and maintenance performance management. *J of Qual in Maint Eng* 9: 333–350
- Lynch RL and Cross KF (1991) Measure up! The Essential Guide to Measuring Business Performance, Mandarin: London.
- Mather D (2005) An introduction to maintenance scorecard. The plant maintenance News letter edition 52, dated 13 April 2005.
- Murthy DNP, Atrons A, and Eccleston JA (2002) Strategic maintenance management. *J of Qual in Maint Eng* 8(4): 287–305.
- Neely AD, Gregory M and Platts K (1995) Performance Measurement System Design – A Literature Review and Research Agenda. *Int J of Oper and Prod Manag* 15(4): 80–116.
- Neely A, Mills J, Platts. K, Richards H, Gregory M, Bourne M, and Kennerley M (2000) Performance measurement system design: developing and testing a process-based approach. *int J of Oper & Prod Manag* 20(10): 1119–1145.
- Parida A (2006) Development of a multi-criteria hierarchical framework for maintenance performance measurement: Concepts, issues and challenges, Doctoral thesis, Luleå University of Technology, 2006:37, ISBN: LTU-DT-06/37-SE, <http://epubl.ltu.se/1402-1544/2006/37/index-en.html>
- Parida A (2007) Role of condition monitoring and performance measurement in asset productivity enhancement. Proceedings of the 19th International Congress COMADEM, 12–14 June 2007, Porto, Portugal, pp 525–531.
- Parida A, Chattopadhyay G and Kumar U (2005) Multi criteria maintenance performance measurement: a conceptual model. Proceedings of the 18th International Congress COMADEM, 31st Aug-2nd Sep 2005, Cranfield, UK, pp 349–356.
- Pun KF and White AS (1996) A performance measurement paradigm for integrating strategy formulation: a review of systems and frameworks. *int J of manag rev* 7: 49–71.

- Raouf A (1994) Improving Capital Productivity through Maintenance. *Int J of Oper & Prod Manag* 14(7):44–52
- Raouf A and Ben-Daya M (1995) Total maintenance management: a systematic approach. *J of Qual in Maint Eng* 1(1):6–14
- Santos SP, Belton V, and Howick S (2002) Adding value to performance measurement by using system dynamics and multicriteria analysis. *Int J of Oper & Prod Manag* 22(11): 1246–1272.
- Singh H, Motwani J and Kumar A (2000) A review and analysis of the state-of-the-art research on productivity measurement. *Ind Manag and Data Syst* 100(5): 234–241
- Tsang AHC (1998) A strategic approach to managing maintenance performance. *J of Qual in Maint Eng* 4(2): 87–94.
- Tsang AHC (2002) Strategic dimensions of maintenance management. *J of Qual in Maint Eng* 8(1): 7–39
- Wireman T (1998) *Developing Performance Indicators for Managing Maintenance*, Industrial Press, Inc, New York.
- Wireman T (2007) How to calculate return on investment for maintenance improvement projects, http://www.reliabilityweb.com/art04/roi_for_maintenanceimprovement_project.pdf, accessed 25 March 2007

Methods and Tools in Maintenance

Failure Mode and Effect Analysis

Mohamed Ben-Daya

4.1 Introduction

Managing risk is a must for any organization. Clause 0.1 of ISO 9004 mentions risk management along with cost and benefit considerations given its importance to the organization and its customers. Clause 5.4.2 also includes risk assessment and mitigation data as necessary inputs for efficient and effective quality planning. Risk management is also important when dealing with equipment failures and their consequence on production, safety and the environment.

For many years failure mode and effect analysis has been used in many sectors to manage risk. Dhillon (1992) traced the history of FMEA back to the early 1950s, when it was used for the design of flight control systems. FMEA emerged as a formal technique in the aerospace and defense industries. It was used on the NASA Apollo missions. The Navy developed an FMEA military standard (MIL STD 1629). FMEA then spread to the American automotive industry in the late 1970s, where car manufacturers start using FMEA in the design of their product development process to deal with their poor reliability and face international competition. FMEA was later adopted by the International Electrochemical Commission in 1985. British Standard BS5760 Part 5 dealing with FMEA is dated 1991. Several books devoted to FMEA appeared in the 1990s (Stamatis, 1995; Palady, 1995; McDermott *et al.* 1996). Many authors adapted FMEA methodology to various industries, such as nuclear power industry (Pinna *et al.* 1998), environmental concerns (Vandenbrande, 1998), software (Goddard, 2000), semi-conductor processing (Whitcomb and Rioux, 1994; Trahan and Pollack, 1999), web-based distributed design (Huang 1999; Wiseman and Denson 1998), healthcare (DeRosier and Stalhandske, 2002; Reiling *et al.* 2003), and this list is by no means exhaustive.

So, what is FMEA? The thought process behind FMEA is implicit in any development process aimed at minimizing risk whether in product development or process analysis. In such an endeavor one has to ask the following logical questions: “What problems could arise?”, “How likely these problems will occur

and how serious they are, if they happen?”, and more importantly “How can these problems be prevented?”

Therefore FMEA is a systematic analysis of potential failure modes aimed at preventing failures. It is intended to be a preventive action process carried out before implementing new or changes in products or processes. An effective FMEA identifies corrective actions required to prevent failures from reaching the customer; and to assure the highest possible yield, quality, and reliability. While Engineers have always analyzed processes and products for potential failures, the FMEA method standardizes the approach and establishes a common language that can be used both, within and between companies.

Some of the benefits of conducting FMEA are:

- Increase customer satisfaction by improving safety and reliability and mitigating the adverse effect of problems before they reach the customer.
- Improve development efficiency in terms of time and cost by solving reliability and manufacturing problems during design stages. The more we move in the development stage, the more rectifying problems becomes more expensive.
- Document, prioritize, and communicate potential risks by making issues explicit to FMEA team members, management, and customers.
- Help reduce the chances of catastrophic failure that can result in injuries and/or adverse effect on the environment.
- Optimize maintenance efforts by suggesting applicable and effective preventive maintenance tasks for potential failure modes. This application of reliability centered maintenance (RCM) will be discussed in more detail in the reliability centered maintenance chapter in this handbook.

The purpose of this chapter is to present FMEA methodology as one of the important tools used in maintenance, especially in reliability centered maintenance, as described in the RCM chapter in this book. Other important tools including root cause analysis, Pareto chart and cause and effect diagram are presented as well.

This chapter is organized as follows: FMEA is defined in the next section and the FMEA process is outlined in Section 4.3. Types of FMEA are described in Section 4.4. Section 4.5 provides some examples of FMEA application in several areas, especially in the service industry.

4.2 FMEA Defined

Failure mode and effect analysis (FMEA) is an engineering technique used to define, identify, and eliminate known and/or potential problems, errors, and so on from the system, design, process, and/or service before they reach the customer (Omdahl, 1988; ASQC, 1983).

It is clear from this definition that FMEA is a systemic methodology intended to perform the following activities:

1. Identify and recognize potential failures including their causes and effects;
2. Evaluate and prioritize identified failure modes since failures are not created equal; and
3. Identify and suggest actions that can eliminate or reduce the chance of the potential failures from occurring.

Ideally, FMEAs are conducted in the product design or process development stages. However, conducting them on existing products and processes may also yield benefits such as in RCM to develop an effective preventive maintenance program.

Section 3.1 of MIL-SRD-1629A makes the following definitions:

- *Failure mode*: the manner by which a failure is observed.
- *Failure cause*: the physical or chemical processes, design defects, quality defects, part misapplication, or other processes which are the basic reason for failure or which initiate the physical process by which deterioration proceeds to failure.
- *Failure effect*: the consequence(s) a failure mode has on the operation, function, or status of an item. Failure effects are classified as local effect, next higher level, and end effect.
- *Local effect*: the consequence(s) a failure mode has on the operation, function, or status of the specific item being analyzed.
- *Next higher level effect*: the consequence(s) a failure mode has on the operation, function, or status of the items in the next higher level of indenture above the indenture level under consideration.
- *End effect*: the consequence(s) a failure mode has on the operation, function, or status of the highest indenture level.
- *Indenture levels*: the item levels which identify or describe relative complexity of assembly or function. The levels progress from the more complex (system) to the simpler (part) division.

Identifying known and potential failure modes is an important task in FMEA. Using data and knowledge of the process or product, each potential failure mode and effect is rated in each of the following three factors:

- *Severity*: the consequence of the failure when it happens;
- *Occurrence*: the probability or frequency of the failure occurring; and
- *Detection*: the probability of the failure being detected before the impact of the effect is realized.

Then these three factors are combined in one number called the risk priority number (RPN) to reflect the priority of the failure modes identified. The risk priority number (RPN) is simply calculated by multiplying the severity rating, times the occurrence probability rating, times the detection probability rating:

$$\text{Risk Priority Number} = \text{Severity} \times \text{Occurrence} \times \text{Detection}$$

These important FMEA tasks are summarized in Figure 4.1.

The failure modes are not created equal and need to be prioritized by ranking them according to the risk priority number from highest to the smallest. A Pareto diagram (Section 4.5.2) can be used to visualize the differences between the various ratings.

The next important FMEA task is to focus limited resources on critical design and/or process issues to improve reliability, quality and safety.

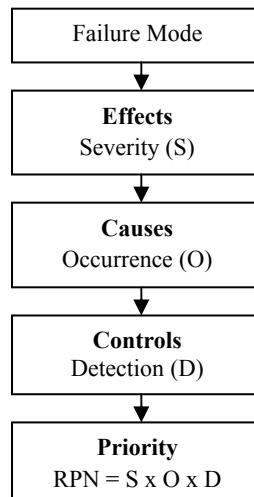


Figure 4.1. Important FMEA tasks

4.3 FMEA Process

Being reactive to quality performance problems by identifying and eliminating the root cause on nonconformities is a common practice. However, a more rewarding challenge is to be ahead of potential problems and designing them out of processes or preventing them from occurring. A typical FMEA process is a proactive methodology that follows the following typical steps:

1. Select a high-risk process.
2. Review the process: this step usually involves a carefully selected team that includes people with various job responsibilities and levels of experiences. The purpose of an FMEA team is to bring a variety of perspectives and experiences to the project.
3. Brainstorm potential failure modes.
4. Identify the root causes of failure modes.
5. List potential effects of each failure mode.
6. Assign severity, occurrence, and detection ratings for each effect.

7. Calculate the risk priority number (RPN) for each effect.
8. Prioritize the failure modes for action using RPN.
9. Take action to eliminate or reduce the high-risk failure modes.
10. Calculate the Resulting RPN as the failure modes are reduced or eliminated as a mean of monitoring the redesigned improved product or process.

Assigning severity occurrence and detection ratings is usually done on a scale from 1 to 10 using tables similar to the ones shown in Tables 4.1–4.3.

A typical way of documenting the FMEA process is by using a matrix similar to the one shown in Table 4.4.

Using RPN analysis to prioritize failure modes has its limitations. In particular:

- Different set of severity, occurrence, and detection may produce the same RPN although the risk implications may be totally different; and
- Severity, occurrence and detection are given the same importance (weight) in RPN calculations.

Generally, there are four types of FMEA: system, design, process, and service FMEA:

- System FMEA focuses on global system functions;
- Design FMEA focuses on components and subsystems;
- Process FMEA focuses on manufacturing and assembly processes; and
- Service FMEA focuses on service functions.

Table 4.1. Typical occurrence evaluation criteria

Probability of Failure	Possible failure rates	Ranking
Very high: failure is almost inevitable	≥ 1 in 2	10
	1 in 3	9
High: repeated failures	1 in 8	8
	1 in 20	7
Moderate: occasional failures	1 in 80	6
	1 in 400	5
	1 in 2,000	4
Low: relatively few failures	1 in 15,000	3
	1 in 150,000	2
Remote: failure is unlikely	≤ 1 in 1,500,000	1

Table 4.2. Typical severity evaluation criteria

Effect	Criteria: severity of effect	Ranking
Hazardous – without warning	Very high severity ranking when a potential failure mode affects safe operation and/or involves noncompliance with regulations without warning	10
Hazardous – with warning	Very high severity ranking when a potential failure mode affects safe operation and/or involves noncompliance with regulations with warning	9
Very high	Product/item inoperable, with loss of primary function	8
High	Product/item operable, but at reduced level of performance. Customer dissatisfied	7
Moderate	Product/item operable, but may cause rework/repair and/or damage to equipment	6
Low	Product/item operable, but may cause slight inconvenience to related operations	5
Very low	Product/item operable, but possesses some defects (aesthetic and otherwise) noticeable to most customers	4
Minor	Product/item operable, but may possess some defects noticeable by discriminating customers	3
Very minor	Product/item operable, but is in noncompliance with company policy	2
None	No effect	1

Table 4.3. Typical detection evaluation criteria

Detection	Criteria: likelihood of detection by design control	Ranking
Absolute uncertainty	Design control will not and/or can not detect a potential cause/mechanism and subsequent failure mode; or there is no design control	10
Very remote	Very remote chance the design control will detect a potential cause/mechanism and subsequent failure mode	9
Remote	Remote chance the design control will detect a potential cause/mechanism and subsequent failure mode	8
Very low	Very low chance the design control will detect a potential cause/mechanism and subsequent failure mode	7
Low	Low chance the design control will detect a potential cause/mechanism and subsequent failure mode	6
Moderate	Moderate chance the design control will detect a potential cause/mechanism and subsequent failure mode	5
Moderately high	Moderately high chance the design control will detect a potential cause/mechanism and subsequent failure mode	4
High	High chance the design control will detect a potential cause/mechanism and subsequent failure mode	3
Very high	Very high chance the design control will detect a potential cause/mechanism and subsequent failure mode	2
Almost certain	Design control will almost certainly detect a potential cause/mechanism and subsequent failure mode	1

Table 4.4. Documentation of FMEA

	Potential failure mode(s)	Potential effects of failure mode	Severity	Potential causes of failure	Occurrence	Present detection systems	Detection	Risk Priority Number (RPN)	Results			
									Recommended action	Severity	Occurrence	Detection

4.4 FMEA Applications

Although FMEA started in the aerospace and automobile industry, it found application in various areas, such as the healthcare industry. With patient safety a priority in healthcare, the technique has seen application in healthcare. Medical devices and medical services such as drug delivery have added FMEA as a means to understand the risks not considered by individual design and process personnel. FMEA allows a team of persons to review the design at key points in product development or medical service and make comments and changes to the design of the product or process well in advance of actually experiencing the failure. The Food and Drug Administration (FDA) has recognized FMEA as a design verification method for Drugs and Medical Devices. Hospitals also use FMEA to prevent the possibility of process errors and mistakes leading to incorrect surgery or medication administration errors. FMEA is now an integral part of many hospitals' continuous improvement program.

4.5 Related Tools

4.5.1 Root Cause Analysis

Finding the real cause of the problem that tends to happen in a repeated fashion and dealing with it rather than simply continuing to deal with the symptoms is called root cause analysis. Root Cause Analysis (RCA) is a step-by-step method used to analyze failures and problems down to their root cause. Every equipment failure happens for a number of reasons. There is a definite progression of actions and consequences that lead to a failure. An RCA investigation traces the cause and effect trail from the end failure back to the root cause in order to determine what happened, why it happened, and more importantly figure out what to do to reduce the likelihood that it will happen again.

The process of analyzing the root cause of failures and acting to eliminate these causes is one of the most powerful tools in improving plant reliability and performance.

Failure investigation process steps are as follows:

1. Problem definition and data gathering

Example of information that should be collected consists of conditions before, during, and after the occurrence; personnel involvement (including actions taken); environmental factors; and other information having relevance to the condition or problem. This is carried out by asking the questions in Table 4.5.

The answers of the questions in Table 4.5 require reviewing records, reports or logs, equipment or installation drawings and documents; conducting interviews with operators, maintenance staff, engineers and plant foremen, and consulting experts regarding possible consequences of corrective actions. We may also need to visit the failed equipment or installation; consulting equipment manufacturer, reviewing computerized information system, *etc.*

Table 4.5. Questions that help define the problem and gather data

Category	Questions
What	• What happened? • What are the symptoms? • What is the complaint? • What went wrong? • What is the undesirable event or behavior?
When	• When did it occur: what date and what time? • During what phase of the production process?
Where	• What plant? • Where did it happen? • What process? • What production stream? • What equipment?
How	• How was the situation before the incident? • What happened during the incident? • How is the situation after the incident? • What is the normal operating condition? • Is there any injury, shutdown, trip, or damage? • How frequent is the problem? • How many other processes, equipments or items affected by this incident?

2. *Control barriers*

Control barriers are administrative or physical aids that are made part of work conditions. They are devices employed to protect employees or equipment and enhance the safety and performance of the machine system. The purpose of checking control barriers in a failure investigation process is to determine if all the control barriers pertaining to the failure under investigation are present and effective. Examples of physical control barriers include conservative design allowance, engineered safety features, fire barriers and seals, ground fault protection, locked doors, valves, breaks, and controls, insulation, redundant system, emergency shutdown system, *etc.*, examples of administrative control barriers include alarms, safety rules and procedures, certification of operators and engineers, methods of communication, policies and procedures, work permits, standards, training and education, *etc.*

3. *Event and causal factor charting*

Event and causal factor charting is an analysis tool whereby events relations, conditions, changes, barriers, and causal factors are charted on a timeline using a standard representation using the symbols shown in Figure 4.2.

4. *Cause and effect analysis*

When the entire occurrence has been charted out, the investigators are in a good position to identify the major contributors to the problem, the causal factors. The diagram will help to show the cause and effect relationship between factors, even if significantly removed from each other in the system.

5. *Root cause identification*

After identifying all causal factors, the team begins the root cause identification. This step generally involves the use of a decision diagram or “fishbone” diagram (see Section 4.5.3). This diagram structures the reasoning process of the investigators by helping them answer questions about why a particular causal factor exists or occurred. For every event there will likely be a number of causal factors. For each causal factor there will likely be a number of root causes.

6. *Corrective actions effectiveness assessment*

The final step of the process is to generate recommendations for corrective action taking into consideration the following questions:

- What can be done to prevent the problem from happening again?
- How will the solution be implemented?
- Who will be responsible for it? and
- What are the risks of implementing the solution?

7. *Report generation*

It is important to report and document the RCA process including a discussion of corrective actions, management and personnel involved. Information of interest to other facilities should also be included in the report.

The study report should include

- Problem definition;
- Event and causal factors chart;
- Cause and effect analysis;
- Root cause(s) of the problem;
- Problem solution; and
- Implementation plan with clear responsibilities and follow-up.

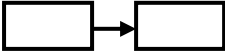
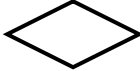
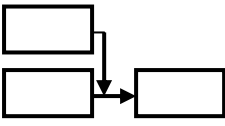
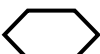
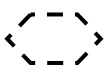
Item	Symbol	Description
Event		An action that occurs during some activity
Primary event		The action directly leading up to or following the primary effect
Undesirable event		An undesirable event (failure, conditions deviation, malfunction, or inappropriate action) that was critical for the situation
Secondary event		An action that impacts the primary event but is not directly involved in the situation
Terminal event		The end point of the analysis
Condition		Circumstances pertinent that may have influenced and/or changed the course of events, or caused the undesirable event
Presumptive event		An action that is assumed because it appears logical in the sequence but cannot be proven
Causal factor		A factor that shaped the outcome of the situation, the root cause of the problem
Presumptive causal factor		A factor that is assumed as it appears to logically affect the outcome
Change		A change in the condition of the situation after an event have occurred
Barrier		Physical or administrative barrier to prevent an unwanted situation
Failed barrier		Physical or administrative barrier that failed to prevent an unwanted situation

Figure 4.2. Standard symbols for factor charting

4.5.2 Pareto Chart

The Pareto chart is one of the seven basic tools of quality control, which include the histogram, Pareto chart, check sheet, control chart, cause-and-effect diagram, flowchart, and scatter diagram. The chart is named after Vilfredo Pareto the Italian economist who noted that 80 % of the income in Italy went to 20 % of the population. The Pareto Principle illustrates the fact that 80 % of the problems stem from 20 % of the causes.

A Pareto Chart is a bar graph made of a series of bars whose heights reflect the frequency of problems or causes. The bars are arranged in descending order of height from left to right. This means the factors represented by the tall bars on the left are relatively more significant than those on the right. This helps sort out the important few from the trivial many so that resources and efforts are focused where we can obtain maximum returns.

A Pareto chart is a helpful tool in any improvement effort and at different levels. It can be used early on to identify which problem should be studied and later on to narrow down which causes of the problem to address first.

To construct a Pareto Chart, one can follow the following steps:

1. *Record the raw data.* List each category and its associated frequency.
2. *Order the data.* Place the category with highest frequency first.
3. *Label the left-hand vertical axis.* Make sure the labels are spaced in equal intervals from 0 to a round number equal to or just larger than the total of all counts.
4. *Label the horizontal axis.* Make the widths of all of the bars the same and label the categories from largest to smallest.
5. *Plot a bar for each category.* The height of each bar should equal the frequency of the corresponding category and their width should be identical.
6. *Find the cumulative counts.* Each category's cumulative count is the count for that category added to the counts for all larger categories preceding it.
7. *Add a cumulative line.* Label the right axis from 0 to 100%, and line up the 100% with the grand total on the left axis. For each category, put a dot as high as the cumulative total and in line with the right edge of that category's bar. Connect all the dots with straight lines.
8. *Analyze the diagram.* Look for the *break point* on the cumulative percent graph that separates the significant few from the trivial many. A clear change in the slope of the graph can help identify the breakpoint.

This procedure is illustrated in Figure 4.3.

4.5.3 Cause and Effect Diagram

A cause-and-effect diagram is a tool that helps identify, sort, and display possible causes of a specific problem or quality characteristic. It graphically illustrates the relationship between a given outcome and all the factors that influence the outcome. This type of diagram is also called a "fishbone diagram" as it resembles the skeleton of a fish. It is also named "Ishikawa diagram" as it was invented by Kaoru Ishikawa.

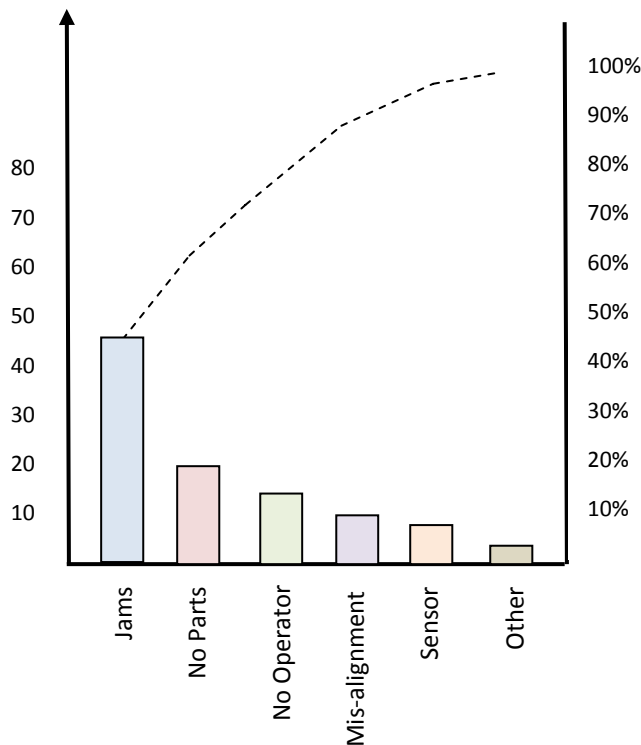


Figure 4.3. Pareto chart

A cause-and-effect diagram is a tool that is helpful for identifying and organizing the causes of a problem such as equipment failure. The structure of the diagram provides a very systematic way of thinking about the causes of a particular problem. Some of the benefits of using this tool are as follows:

- Identifies the *root causes* of a problem using a structured approach;
- Promotes group participation and utilizes group knowledge of the process;
- Uses an orderly, easy-to-read format to diagram cause-and-effect relationships;
- Increases knowledge of the process by helping everyone to learn more about the factors at work and how they relate to the problem;
- Identifies areas where data should be collected for further study, if needed; and
- Constructs a pictorial display of a list of causes organized in different categories to show their relationship to a particular problem or effect.

Figure 4.4 shows the basic layout of a cause-and-effect diagram. Notice that the diagram has a cause side and an effect side. The steps for constructing and analyzing a Cause-and-Effect Diagram are outlined below:

1. Identify and clearly define the problem or effect to be analyzed. It is a good practice to develop an Operational Definition of the effect to ensure that it is clearly understood by all team members.
2. Draw a horizontal arrow pointing to the right. This is the spine. To the right of the arrow, write a brief description of the effect or problem to be analyzed
3. Identify the main causes contributing to the effect being studied. These are the labels for the major branches of the diagram and become categories under which to list the many causes related to those categories. Some commonly used categories are as follows:
 - Methods, materials, machinery, and people (3Ms and P);
 - Policies, procedure, people, and plant (4Ps); and
 - Another possible significant fifth factor is the environment.
4. For each major branch or category, identify other specific factors which may be the causes of the effect under that category. Identify as many causes or factors as possible and attach them as sub-branches of the major branches.
5. Identify more detailed levels of causes and continue organizing them under related causes or categories.
6. Analyze the diagram. Analysis helps you identify causes that warrant further investigation. Since cause-and-effect diagrams identify only possible causes, you may want to use a Pareto Chart to help determine the causes to focus on first.

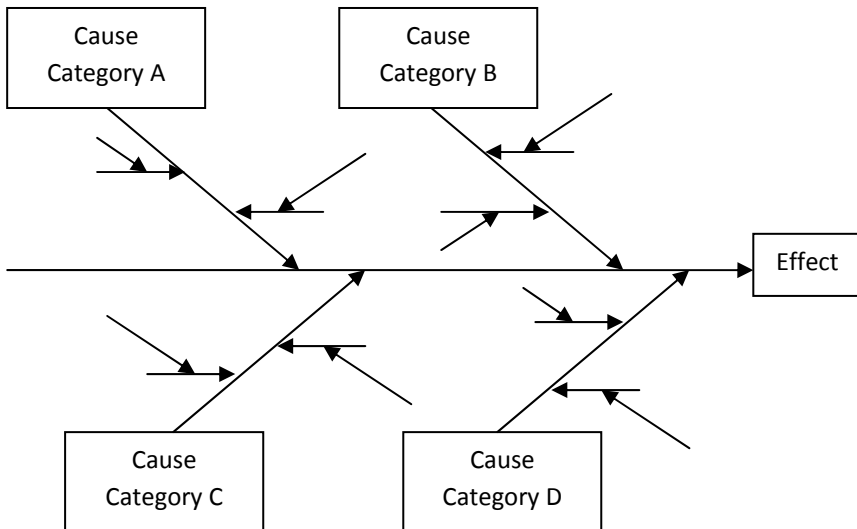


Figure 4.4. Cause and effect diagram

References

- American Society for Quality Control Statistics Division (1983) Glossary and tables for statistical quality control. 2nd edn. ASQC Quality Press, Milwaukee.
- British Standard BS5760 Part 5 (1991) Guide to Failure Modes, Effects and Criticality Analysis (FMEA and FMECA). British Standards Publishing Ltd., London, UK.
- DeRosier J and Stalhandske E (2002) The Healthcare Failure Mode Effect Analysis process, Version August 2002. United States Department of Veteran Affairs National Center for Patient Safety.
- Dhillon, BS (1992) Failure Modes and Effects Analysis – Bibliography. *Microelectroni Reliab* 32:719–731.
- Ford Motor Company (1992) FMEA Handbook. Ford Motor Company, Engineering Materials and Standards, Dearborn, Michigan.
- Goddard PL (2000) Software FMEA Techniques. *Proceedings of the 2000 IEEE Annual Reliability and Maintainability Symposium*, pp 118–123.
- Huang GQ, Nie M, *et al.* (1999) Web-based Mode and Effect Analysis (FMEA). *Comput and Ind Eng* 37:177–180.
- McDermott RE, Mikulak RJ, Beauregard MR (1996) The Basics of FMEA. OR: Productivity, Inc, Portland.
- Omdahl TP, ed. (1988) Reliability, availability, and maintainability dictionary. ASQC Quality Press, Milwaukee.
- McDermott RE, Mikulak RJ, Beauregard MR (1996) The basics of FMEA. Quality Resources, New York, NY.
- Palady P (1995) Failure Modes and Effects Analysis: Predicting & Preventing Problems before they Occur. PT Publications, West Palm Beach, FL.
- Pinna T, Caporali R, Cambi G, Buragazzi L, Poucet A, Porfiri MT (1998) Failure mode and Effect Analysis on International Thermonuclear Experimental Reactor (ITER) Heat Transfer. *Fusion Eng and Des* 42: 431–436.
- Reiling GJ, Knutzen BL, Stoecklein M (2003) FMEA—the cure for medical errors. *Qualit Prog* 36(8):67–71.
- Stamatis, DH (1995) Failure Mode and Effect Analysis: FMEA from Theory to Execution. ASQC Quality Press, Milwaukee, WI.
- Trahan R and Pollack A (1999) Using an Inverted FMEA to Manage Change and Reduce Risk in a FAB. *Proceedings of the 1999 IEEE International Symposium on Semiconductor Manufacturing Conference*, pp 15–18.
- United States Department of Defense (1974/80/84) Procedures for performing a Failure Mode, Effects and Criticality Analysis. US MIL STD 1629 (1974), US MIL-STD 1629A (1980), US MIL-STD 1629A/Notice 2 (1984), Washington, DC.
- Vandenbrande WW (1998) How to use FMEA to reduce the Size of your Quality Toolbox. *Qualit Prog* 31:97–100.
- Whitcomb R, Rioux M (1994) Failure Modes and Effects Analysis (FMEA) System Development in a Semiconductor Manufacturing Environment. *Proceedings of the 1994 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop*, pp 136–139.
- Wiseman M and Denson W (1998) Systems Monitors Reliability. *Pulp and Pap Can* 99:32–34.

Failure Statistics

Mohamed Ben-Daya

3.1 Introduction

Probability and statistics are indispensable tools in reliability maintenance studies. Although excellent texts exist in these areas, an introduction containing essential concepts is included to make the handbook self-contained.

This chapter is organized as follows. The next section provides an introduction to basic probability concepts. The third section introduces the reliability and failure rate function. Commonly used probability distributions are presented in Section 3.5. Finally, Section 3.6 deals with types of data and parameter estimation.

3.2 Introduction to Probability

3.2.1 Sample Spaces and Events

The time to failure of a device can vary for the same item and for similar items. This is an example of a random experiment that can be defined as follows:

A random experiment is an experiment that can result in different outcomes, even though it is repeated in the same manner every time.

The outcome of a statistical experiment is not predictable in advance. However the entire set of possible outcomes is known and is defined as follows:

The set of all outcomes of a statistical experiment is known as the *sample space* and is denoted by S .

Suppose that the experiment consists of tossing a coin. Then there are two possible outcomes, namely heads and tails. Hence

$$S = \{H, T\}$$

Suppose that the experiment consists of recording the lifetime T of some equipment; then the sample space S consists of all nonnegative real numbers. Thus

$$S = \{T \mid T \geq 0\}$$

In a statistical experiment one may be interested in the occurrence of particular points in the sample space S . For example, consider the experiment which consists of tossing a die. A subset of interest might be that obtained by considering only odd outcomes. Such subsets are called events:

An *event* is any subset of the sample space S .

Consider the sample space $S = \{t \mid t \geq 0\}$, where t is the life of some equipment. Let $E = [3, 5]$, then E is the event that the lifetime of the equipment is between 3 and 5 years.

The sample space S itself is an event containing all possible outcomes of the experiment. Given any set E , a closely related subset to E is defined as follows:

The *complement* of an event E is the set of all points that are not in E and is denoted by $\bar{E}$.

Consider the sample space $S = \{1, 2, 3, 4, 5, 6\}$ consisting of all possible outcomes obtained when tossing a die. If $E = \{1, 3, 5\}$, then its complement $\bar{E} = \{2, 4, 6\}$.

3.2.2 Definition of Probability

The probability of an outcome can be interpreted as the limiting value of the proportion of times the outcome occurs in n repetitions of the random experiment as n increases beyond all bounds (Montgomery and Runger (1999)). Consider a statistical experiment with sample space S and let E be an event of S , then

The probability of event E , $P(E)$, is a number assigned to each member of a collection of events from a random experiment that satisfies the following conditions:

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. If $E_1, E_2, \dots, E_n$ are mutually exclusive events, then

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = P(E_1) + P(E_2) + \dots + P(E_n)$$

Consider the experiment of flipping a coin. If the events $E_1 = \{H\}$ and $E_2 = \{T\}$ are equally likely to occur then $P(\{H\}) = P(\{T\}) = 1/2$

3.2.3 Probability Rules

1. Let $\bar{E}$ be the complement of E , that is the set of all points in the sample space S which are not in E then since E and $\bar{E}$ are mutually exclusive and $E \cup \bar{E} = S$, we have

$$1 = P(S) = P(E \cup \bar{E}) = P(E) + P(\bar{E})$$

Thus

$$P(\bar{E}) = 1 - P(E)$$

This rule says that the probability that an event does not occur is equal to one minus the probability that it does occur.

2. This rule provides a formula for the probability of the union of two events E and F in S , $E \cup F$:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

Note that $P(E \cap F)$ is subtracted because any point in $E \cap F$ is counted twice in $P(E) + P(F)$. If E and F are mutually exclusive, then $P(E \cap F) = \Phi$, the empty set. Hence $P(E \cup F) = P(E) + P(F)$. This result can be obtained from the third condition in the definition of probability.

3.2.4 Conditional Probabilities

Suppose that 10 coins are numbered one through 10 and mixed up so that each coin is equally likely to be drawn. One coin is drawn from the 10 coins. Suppose that we know that the number on the drawn coin is at least 7. Given this information, what is the probability that it is 9?

Knowing that the number on the drawn coin is at least 7, then it follows that the possible outcomes of this experiment are 7, 8, 9 or 10. These outcomes have the same (conditional) probability of occurring, namely $1/4$ while the probability of the other outcomes (1, 2, 3, 4, 5, 6) is 0. Notice that the probability of drawing the number 9 without the information that it is at least 7 would have been $1/10$. The probability of obtaining number 9 given that the number on the coin drawn is at least 7 is called conditional probability.

A formal definition of conditional probability is as follows:

$$P(B | A) = \frac{P(A \cap B)}{P(A)}, \quad \text{if } P(A) > 0.$$

The rationale behind this formula is that if we know that A occurs, then for B to occur it is necessary for the outcome to be a point in both A and B , that is, the outcome belongs to $A \cap B$. Knowing that A has occurred reduces the sample space of the experiment to A . Hence the probability of $A \cap B$ occurring is equal to the probability of $A \cap B$ relative to the probability of A .

Example 3.1: A coin is tossed twice. What is the conditional probability that both outcomes are heads given that at least one of them is heads? Assume that the sample space $S = \{HH, HT, TH, TT\}$, and all the outcomes are equally likely.

Let B denote the event that both tosses come heads, and A the event that at least one of the tosses comes heads, then the conditional probability is given by

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{P(\{HH\})}{P(\{HH, HT, TH\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

There are situations in which the occurrence of A has no impact on the occurrence of B , that is the occurrence of B is independent of the occurrence of A . These two events are *independent*. The following is a formal definition of independence of two events A and B :

Two events A and B are said to be independent if $P(A \cap B) = P(A)P(B)$.

This definition implies that $P(B|A) = P(B)$ and $P(A|B) = P(A)$. The concept of independence is very important and plays a vital role in many applications of probability and statistics.

Example 3.2: Consider the experiment of tossing two dice. Let A be the event that the first die equals 4 and B the event that the sum of the dice is 9. Are A and B independent?

$$P(A \cap B) = \frac{P(\{4,5\})}{P(A)} = \frac{1}{36}$$

$$P(A)P(B) = \frac{1}{6} \frac{1}{6} = \frac{1}{36}.$$

Hence events A and B are independent.

3.2.5 Random Variables

Very often we are more interested in some function of the outcome of an experiment rather than the outcome itself. Each element in the sample space is assigned a numerical value. These values are random quantities determined by the outcome of the experiment. The functions that assign to each element of the sample space some value are called random variables. A formal definition is as follows:

A random variable is a function that associates a real number with each element of the sample space.

The value of a random variable is determined by the outcome of the experiment, therefore probabilities may be assigned to the values of the random variable.

The following examples will be used to clarify the concept of random variables.

Example 3.3: Suppose that the experiment consists of flipping two coins. Let the random variable X be the number of tails appearing. Then X can take the values 0, 1, and 2. The corresponding probabilities are given by:

$$\begin{aligned}
 P(X = 0) &= P(\{HH\}) = \frac{1}{4} \\
 P(X = 1) &= P(\{HT, TH\}) = \frac{2}{4} \\
 P(X = 2) &= P(\{TT\}) = \frac{1}{4}
 \end{aligned}$$

Random variables are classified as discrete or continuous according to the values they can assume. These two classes of random variables are defined as follows:

A random variable is said to be *discrete* if its set of possible values is countable.

A random variable is said to be *continuous* if its set of possible values is uncountable.

Continuous random variables take on values on a continuous scale.

3.3 Probability Distributions

For a discrete random variable X , we define the probability mass function $f(x)$ of X by:

The set of ordered pairs $(x, f(x))$ is a probability mass function of the discrete random variable X if for each outcome x :

1. $f(x) \geq 0$;
2. $\sum_x f(x) = 1$; and
3. $P(X = x) = f(x)$.

Example 3.4: A shipment of nine spare parts to a plant warehouse contains three defective parts. Assume that two of the nine parts are randomly issued to the maintenance department. Let the random variable X be the number of defectives issued to the maintenance department. What is the probability mass function of X ?

Note that X can assume the values 0, 1, and 2. Hence

$$\begin{aligned}
 f(0) &= P(X = 0) = \frac{\binom{6}{2}\binom{3}{0}}{\binom{9}{2}} = \frac{15}{36} \\
 f(1) &= P(X = 1) = \frac{\binom{6}{1}\binom{3}{1}}{\binom{9}{2}} = \frac{18}{36} \\
 f(2) &= P(X = 2) = \frac{\binom{6}{0}\binom{3}{2}}{\binom{9}{2}} = \frac{3}{36}
 \end{aligned}$$

Thus the probability mass function is given by the following table:

x	0	1	2
$f(x)$	$15/36$	$18/36$	$3/36$

The cumulative distribution of the random variable X is defined as follows:

The cumulative distribution $F(x)$ of a discrete random variable X with probability mass function $f(x)$ is given by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \quad \text{for } -\infty \leq x \leq \infty.$$

Example 3.5: For Example 3.4, the cumulative distribution of X is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{18}{36} & \text{for } 0 \leq x < 1 \\ \frac{33}{36} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

For a continuous random variable X , we define the *probability density function* $f(x)$ of X by:

The set of ordered pairs $(x, f(x))$ is a probability density function of the continuous random variable X if:

1. $f(x) \geq 0$, for all $x \in R$;
2. $\int_{-\infty}^{\infty} f(x) dx = 1$; and
3. $P(a < x < b) = \int_a^b f(x) dx$.

Example 3.6: Consider a random variable X which takes value on the interval $[a, b]$ such that the probability that X is any particular subinterval of $[a, b]$ equals the length of that subinterval. Find the probability density function of X .

The random variable X is known as the *uniform* random variable and its density is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

The cumulative distribution of the random variable X is defined as follows:

The cumulative distribution $F(x)$ of a continuous random variable X with probability density function $f(x)$ is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad \text{for } -\infty < x < \infty$$

Example 3.7: For Example 3.6, the cumulative distribution of X is given by

$$f(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x > b \end{cases}$$

The cumulative distribution $F(x)$ satisfies the following properties:

1. $0 \leq F(x) \leq 1$;
2. If $a < b$; $F(a) < F(b)$; and
3. $F(-\infty) = 0$ and $F(\infty) = 1$.

3.4 Reliability and Failure Rate Functions

3.4.1 Introduction

In this section, we introduce the reliability and failure rate functions and mean time to failure. We motivate this discussion using the following example which helps the reader understand the concepts being introduced.

Example 3.8: A maintenance department is keeping history record about the failure pattern of 100 identical electronic components in common use by the electrical section. This data is summarized in Table 3.1 where time is in number of years.

Table 3.1. Data for Example 3.8

Time	1	2	3	4	5	6	7	8	9	10	>10
Number of failures	22	16	12	10	8	7	5	4	4	3	9

Consider estimating the probability distribution associated with the failure of one of these components chosen randomly. Let T be the random variable defining the lifetime of the component, which is the time the component will operate before failure. Using the above data we can estimate the cumulative distribution function $F(t)$ of the random variable T .

Recall from Section 3.3 that $F(t) = P(T \leq t)$. The estimation of $F(t)$ from the above data require finding the cumulative number of failures and the proportion this number represents each year with respect of the total number of failures. This information is given in Table 3.2.

Table 3.2. Cumulative number of failures and frequency of failures

Time	1	2	3	4	5	6	7	8	9	10	>10
Number of failures	22	16	12	10	8	7	5	4	4	3	9
Cumulative number	22	38	50	60	68	75	80	84	88	91	100
Frequency of failures	.22	.38	.50	.60	.68	.75	.80	.84	.88	.91	1.0

The proportion of total are estimates of the cumulative distribution function $F(t)$, for $t=1, 2, \dots$. For example, $F(5) = 68/100 = 0.68$

3.4.2 Reliability Function

Let T be a random variable defining the lifetime of a component with distribution function $F(t)$. If $F(t)$ is a differentiable function, then the probability density function of T is given by

$$f(t) = \frac{dF(t)}{dt}$$

The reliability function $R(t)$ of the component is given by:

$$R(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

It is the probability that the component will operate after time t , sometimes called survival probability.

3.4.3 Failure Rate Function

The failure rate of a system during the interval $[t, t+\Delta t]$ is the rate at which failures occur in the given interval. It can be defined as follows:

The failure rate of a system during the interval $[t, t+\Delta t]$ is the probability that a failure per unit time occurs in the interval, given that a failure has not occurred prior to t , the beginning of the interval.

The conditional probability of failure during the interval $[t, t+\Delta t]$ given that a failure has not occurred prior to t is given by

$$\frac{\int_t^{t+\Delta t} f(t)dt}{\int_t^{\infty} f(t)dt} = \frac{F(t+\Delta t) - F(t)}{R(t)}$$

To find the conditional probability per unit time, we divide by Δt . Thus the failure rate is given by

$$\frac{F(t+\Delta t) - F(t)}{\Delta t R(t)}$$

The failure rate function or hazard function is defined as the limit of the failure rate as the interval approaches zero. Hence

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{F(t+\Delta t) - F(t)}{\Delta t R(t)} = \frac{1}{R(t)} \left(\lim_{\Delta t \rightarrow 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} \right) = \frac{1}{R(t)} \frac{dF(t)}{dt} = \frac{f(t)}{R(t)}$$

Therefore the hazard function $h(t)$ is given by

$$h(t) = \frac{f(t)}{R(t)}$$

The failure rate or hazard function can also be derived as follows. Consider the following conditional probability:

$$P(t < T \leq t + s \mid T > t)$$

Given that the component survived past time t , this probability is the conditional probability that the component will fail between t and $t+s$.

Recall from Section 3.2.4 that for any events A and B:

$$P(A \mid B) = \frac{A \cap B}{P(B)}$$

Let $A = \{ t < T \leq t+s \}$ and $B = \{ T > t \}$. Note that $A \subset B$. Hence $A \cap B = A$. Therefore

$$P(t < T \leq t + s \mid T > t) = \frac{P(t < T < t + s)}{P(T > t)} = \frac{F(t + s) - F(t)}{R(t)}$$

Dividing by s and take the limit as $s \rightarrow 0$, we obtain

$$\lim_{s \rightarrow 0} \frac{F(t + s) - F(t)}{s R(t)} = \frac{1}{R(t)} \lim_{s \rightarrow 0} \frac{F(t + s) - F(t)}{s} = \frac{f(t)}{R(t)}$$

This ratio is nothing but the *failure rate or hazard function* $h(t)$

The derivation leading to the expression of $h(t)$ helps also understand the meaning of this important function. The hazard function is the rate of change of the conditional probability of failure at time t . It measures the likelihood that a component that has operated up until time t fails in the next instant of time.

3.4.4 Mean Time Between Failure (MTBF)

The mean time between failure (MTBF) can be obtained by finding the expected value of the random variable T , time to failure. Hence

$$MTBF = E(T) = \int_0^{\infty} t f(t) dt$$

where T is a continuous random variable.

It is worth noting that there is an alternative way for computing the expected value, namely:

$$MTBF = E(T) = \int_0^{\infty} R(t) dt$$

Example 3.9: The reliability function of an electric fan is given by $R(t) = e^{-0.0008t}$. What is the MTBF for this fan?

$$MTBF = E(T) = \int_0^{\infty} R(t) dt = \int_0^{\infty} e^{-0.0008t} dt = -\frac{1}{0.0008} [0 - 1] = 1250$$

3.5 Commonly Used Distributions

In this section, we will concentrate on the most commonly used and most widely applicable distributions for life data analysis, having applications in reliability analysis and maintenance studies, as outlined in the following sections.

3.5.1 The Binomial Distribution

The binomial distribution is a discrete distribution. It arises in cases where many independent trials can result in either a success or a failure and we are interested in finding the probability of having x successes in n such trials.

If we let

- p = the probability of success;
- q = the probability of failure, where $q = 1 - p$;
- n = the number of independent trials;
- x = the number of successes in n trials; and
- $f(x)$ = the probability of x successes

then the probability mass function (pmf) of the binomial distribution is graphed in Figure 3.1 and its expression is given by

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad \text{for } 0 \leq p \leq 1; \quad x = 1, 2, \dots, n \quad (3.1)$$

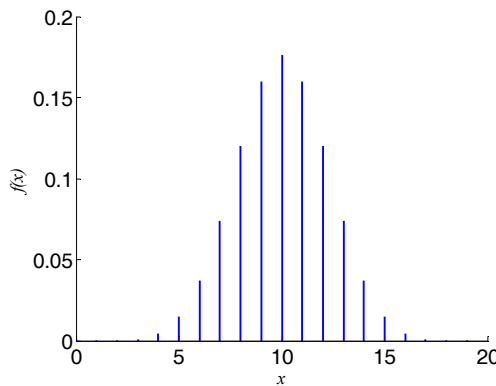


Figure 3.1. Graph of the pmf of the binomial distribution with $p = 0.5$, $n = 20$

The mean of the binomial distribution is $\mu = np$ and variance npq .

Example 3.10: The probability that a certain kind of component will survive a given shock test is 0.75. What is the probability that exactly one out of four components tested will survive the shock test?

This probability can be obtained using Equation 3.1 by letting $n=4, x=1, p=0.75$ and $q=0.25$. Hence

$$f(x) = \binom{4}{1} 0.25^1 0.75^{4-1} = \frac{9}{128}$$

3.5.2 The Poisson Distribution

The Poisson distribution is a discrete distribution that has many applications. It can be used when one is interested in finding the probability of having x failures during a certain period of interest.

If we let

λ = the rate of success;

x = the number of failures during time t ;

$f(x)$ = the probability of x successes

then the pmf of the Poisson distribution is given by

$$f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}; \quad x = 0, 1, 2, \dots \quad (3.2)$$

Its graph is shown in Figure 3.2. The mean and variance of the Poisson distribution are both given by λt .

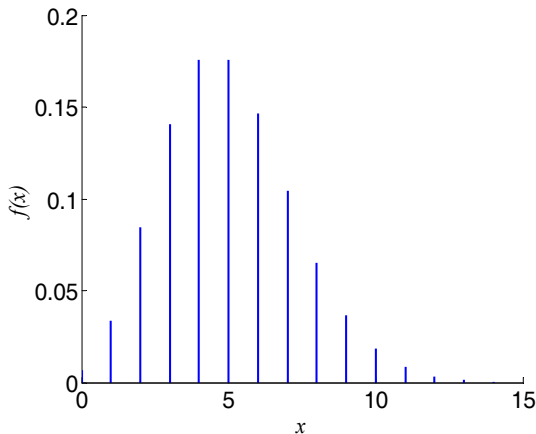


Figure 3.2. Graph of the pmf of the Poisson distribution with $\lambda t = 0.5$

Example 3.11: Suppose that a system contains a certain type of component whose rate of failure is five per year. What is the probability that two components will fail during the first year in the system.

This probability is given by Equation 3.2 by letting $\lambda = 5$, $t = 1$ and $x = 2$. Thus

$$f(2) = \frac{(5 \times 1)^2 \times e^{-5 \times 1}}{2!} = 0.0842$$

3.5.3 The Normal Distribution

The normal distribution is a continuous distribution that has wide applications. It takes the well known bell shape and is symmetrical about its mean value (see Figure 3.3).

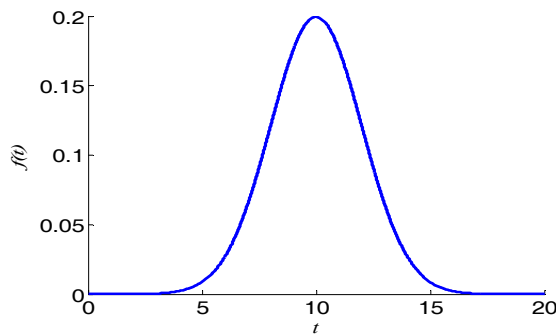


Figure 3.3. Graph of the pdf of the normal distribution with $\mu = 10$ and $\sigma = 2$

Its probability density function is given by

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \quad (3.3)$$

and its cumulative distribution is given by

$$F(t) = \int_{-\infty}^t \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\tau-\mu}{\sigma}\right)^2} d\tau \quad (3.4)$$

There is no closed form for this integral; however tables for the standard normal distribution ($\mu = 0$ and $\sigma = 1$) are readily available and can be used to find the probability for any normal distribution.

The probability density function of the standard normal distribution is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (3.5)$$

and its cumulative distribution is given by

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau \quad (3.6)$$

For a normally distributed random variable T with mean μ and variance σ , its CDF can be expressed in terms of the standard cumulative normal distribution as follows:

$$F(t) = P(T \leq t) = P\left(z \leq \frac{t - \mu}{\sigma}\right) = \Phi\left(\frac{t - \mu}{\sigma}\right) \quad (3.7)$$

Therefore $F(t)$ can be evaluated for any value of t using standard normal tables that are widely available.

The failure rate function, $h(t)$, corresponding to a normal distribution is a monotonically increasing function of t . Graphs of the normal CDF and normal failure rate functions are given in Figures 3.4 and 3.5, respectively.

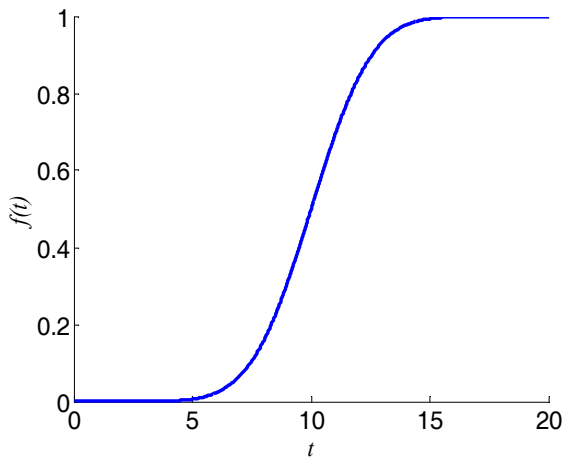


Figure 3.4. Graph of the CDF of the normal distribution with $\mu = 10$ and $\sigma = 2$

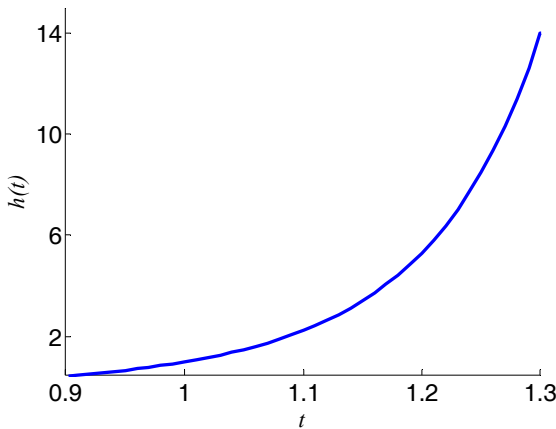


Figure 3.5. Graph of the normal failure rate function with $\mu = 1$ and $\sigma = 0.2$

Example 3.12: A component has normally distributed failure time with $\mu = 20$ and $\sigma = 2$. Find the reliability of the component at 18 time units.

$$R(t) = 1 - F(t) = 1 - \Phi\left(\frac{t - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{18 - 20}{2}\right) = 0.8413$$

3.5.4 The Lognormal Distribution

The probability density function of the lognormal is given by

$$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2} \quad (3.8)$$

where μ and σ are the parameters of the distribution with $\sigma > 0$. Graph of the lognormal pdf is shown in Figure 3.6.

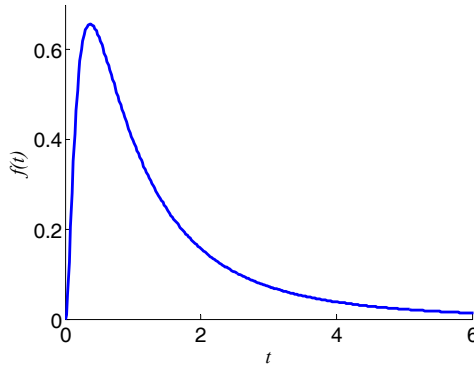


Figure 3.6. Graph of the pdf of the lognormal distribution with $\mu = 0$ and $\sigma = 1$

Note that if a random variable X is defined as $X = \ln T$, where T is lognormally distributed with parameters μ and σ , then X is normally distributed with mean μ and standard deviation σ . This relationship can be exploited to make use of the standard normal in lognormal distribution computations.

The mean and variance of the lognormal are given by

$$\text{Mean} = e^{\mu + \frac{\sigma^2}{2}}, \text{ and}$$

$$\text{Variance} = e^{2\mu + 2\sigma^2} (e^{\sigma^2} - 1)$$

The CDF of the lognormal is given by

$$F(t) = \int_0^t \frac{1}{\sigma \tau \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln \tau - \mu}{\sigma}\right)^2} d\tau \quad (3.9)$$

It can be related to the standard normal as follows:

$$F(t) = P(T \leq t) = P\left(z \leq \frac{\ln t - \mu}{\sigma}\right) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad (3.10)$$

The reliability function is given by

$$R(t) = P(T > t) = P\left(z > \frac{\ln t - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad (3.11)$$

Thus the failure rate function is given by

$$h(t) = \frac{f(t)}{R(t)} = \frac{\phi\left(\frac{\ln t - \mu}{\sigma}\right)}{t \sigma \left[1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)\right]}, \quad (3.12)$$

where ϕ and Φ are the pdf and CDF of the standard normal, respectively.

Graphs of the CDF and failure rate of the lognormal are shown in Figures 3.7 and 3.8, respectively.

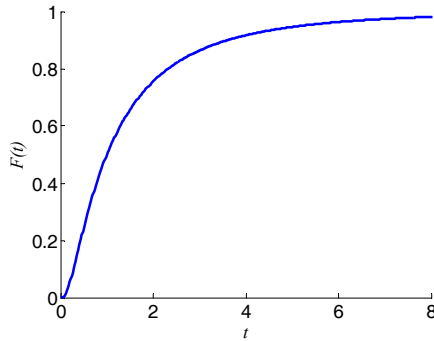


Figure 3.7. Graph of the CDF of the lognormal distribution with $\mu = 0$ and $\sigma = 1$

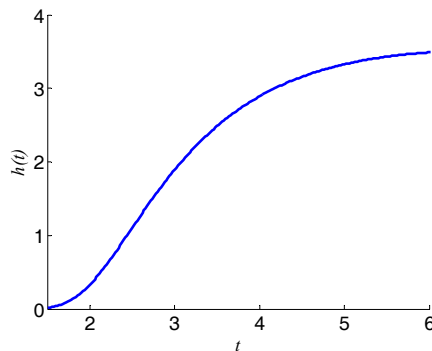


Figure 3.8. Graph of the lognormal failure rate function with $\mu = 1$ and $\sigma = 0.2$

Example 3.13: The failure time of a device follows a log normal distribution with $\mu = 4$ and $\sigma = 1$. Find the reliability and failure rate of the device at $t = 100$.

Using Equation 3.11, $R(100)$ is given by

$$R(100) = 1 - \Phi\left(\frac{\ln 100 - 4}{1}\right) = 0.2725,$$

and using Equation 3.12

$$h(100) = \frac{\phi\left(\frac{\ln 100 - 4}{1}\right)}{100 \left[1 - \Phi\left(\frac{\ln 100 - 4}{1}\right)\right]} = 0.012 \text{ failures/unit time}$$

3.5.5 The Exponential Distribution

The exponential distribution is a continuous distribution that has wide applications. It can be used in reliability as a model of the time to failure of a component.

The probability density function of the exponential distribution is given by

$$f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}}, \quad t \geq 0 \quad (3.13)$$

where $\theta > 0$ is a constant.

The mean and variance of the exponential distribution μ and σ^2 are given by

$$\mu = \theta$$

$$\sigma^2 = \theta^2$$

When the exponential distribution is used as a model of the time to failure of some component or system then θ is the mean time to failure. Also the failure rate is constant and equal to $1/\theta$.

Example 3.14: Suppose that a component has useful life that is satisfactorily modeled by an exponential distribution with mean $\theta = 1000$. What is the probability that this component would fail before 2000?

This probability is given by

$$P(X \leq 2000) = \int_0^{2000} 0.001 e^{-0.001t} dt = 1 - e^{-2} = 0.8467$$

There is an interesting relationship between the exponential and Poisson distribution. Suppose that we use the Poisson distribution as a model of the number of failures of some component in the time interval $(0, t]$; then, using Equation 3.2, the probability of no failure occurring in $(0, t]$ is given by

$$f(0) = e^{-\frac{t}{\theta}}, \quad t \geq 0$$

Let X be the random variable denoting the time to the first failure. The probability that the length of time until the first failure exceeds x is the same as the probability that no Poisson failures will occur in x . The latter is given by $e^{-x/\theta}$ as shown above. Consequently, $P(X > x) = e^{-x/\theta}$. Hence, the cumulative distribution function of the random variable X is given by

$$P(X \leq x) = 1 - e^{-\frac{x}{\theta}},$$

which is the cumulative distribution function of the exponential distribution.

3.5.6 The Weibull Distribution

The Weibull distribution is one of the most widely used lifetime distributions in reliability and maintenance engineering. It is a versatile distribution that can take different shapes. Depending on the value of the shape parameter, β , its failure rate function can be decreasing, constant, or increasing. As such it can be used to model the failure behavior of several real life systems.

The probability density function of the three-parameter Weibull distribution is given by

$$f(t) = \frac{\beta}{\theta} \left(\frac{t - \delta}{\theta} \right)^{\beta-1} e^{-\left(\frac{t - \delta}{\theta} \right)^\beta}, \quad (3.14)$$

where $t \geq 0$, $\delta, \beta, \theta > 0$, and θ is the scale parameter, β is the shape parameter, and δ is the location parameter.

The probability density function of the two-parameter Weibull distribution is given by

$$f(t) = \frac{\beta}{\theta} \left(\frac{t}{\theta} \right)^{\beta-1} e^{-\left(\frac{t}{\theta} \right)^\beta} \quad (3.15)$$

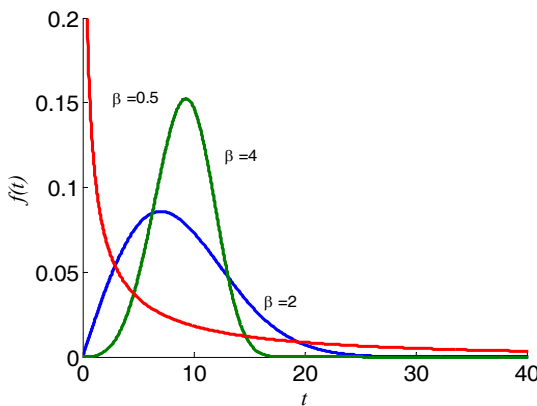


Figure 3.9. Graph of the pdf of Weibull distribution ($\theta=10$)

The graph of the probability density function of the two-parameter Weibull distribution is shown in Figure 3.9 for various values of the shape parameter. Its cumulative distribution function is given by

$$F(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^\beta} \quad (3.16)$$

Its reliability function is given by

$$R(t) = e^{-\left(\frac{t}{\theta}\right)^\beta} \quad (3.17)$$

Graph of the Weibull reliability function is shown in Figure 3.10.

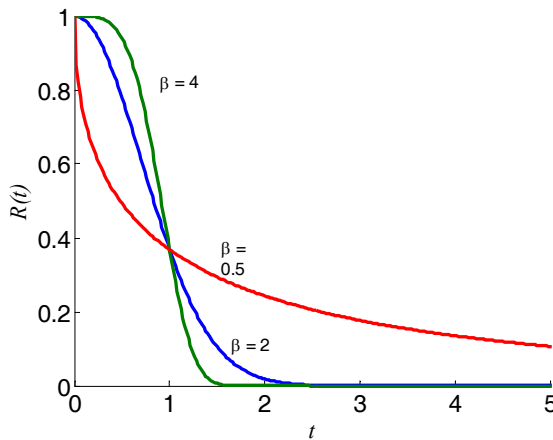


Figure 3.10. Graph of the Weibull reliability function ($\theta=10$)

The mean time to failure is given by

$$MTTF = \theta \Gamma\left(1 + \frac{1}{\beta}\right), \quad (3.18)$$

where $\Gamma(\cdot)$ is the gamma function defined as $\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$.

The corresponding failure rate function is given by

$$h(t) = \frac{\beta}{\theta} \left(\frac{t}{\theta}\right)^{\beta-1} \quad (3.19)$$

Graph of the Weibull failure rate function is shown in Figure 3.11.

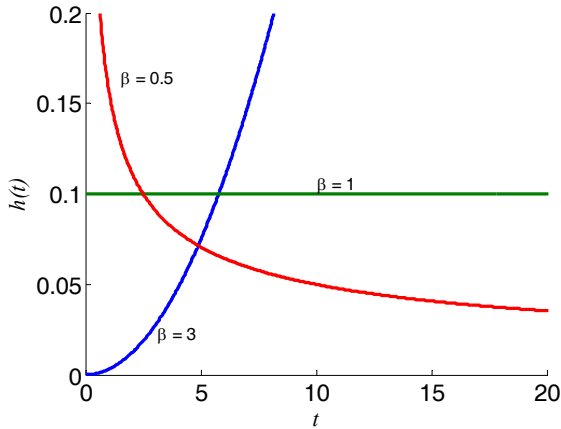


Figure 3.11. Graph of the Weibull hazard function ($\theta=10$)

Example 3.15: The time to failure of an electronic component follows a two-parameter Weibull distribution with $\beta = 0.5$ and $\theta = 800$ h. What is the mean time to failure and what is the fraction of components expected to survive 3200 h?

$$MTTF = \theta \Gamma\left(1 + \frac{1}{\beta}\right) = 800 \Gamma\left(1 + \frac{1}{0.5}\right) = 16000 \text{ h}$$

The fraction of components expected to survive 3200 h is given by

$$R(t) = e^{-\left(\frac{t}{\theta}\right)^{\beta}} = e^{-\left(\frac{3200}{800}\right)^{0.5}} = 0.135.$$

This means that 13.5% of the components will survive 3200 h.

3.6 Failure Statistics

3.6.1 Types of Data

Most failure data can be classified into two categories: complete data and censored data.

Complete data means that our data set is composed of the times-to-failure of all units in our sample. For example, if we tested five devices on a testing stand and all of them failed, then the recorded times-to-failure would provide complete information as to the time of each failure in the sample.

Censored data arises in situation where some units in the sample may not have failed or the exact times-to-failure of all the units are not known at the time of

failure data analysis. There are three types of censored data, right censored (also called suspended data), interval censored, and left censored.

Right censored data are composed of units that did not fail. For example, if we tested ten devices and only seven had failed by the end of the test, we would have suspended data (or right censored data) for the three devices that did not fail. The term "right censored" implies that the time-to-failure is to the right of our data point. In other words, if the three units were to continue operating, the failure would occur to the right of our data point on the time scale.

Interval censored data reflect uncertainty as to the exact times the devices failed within an interval. For example, if we inspect a machine at 1000 and find it operating and then inspect it at 1200 and find it not operating, then we will not know the exact time of the failure. All what we know is that the failure occurred in the interval [1000, 1200].

Left censored data also reflect uncertainty as to the exact time of failure. However, in this case, a failure time is only known to be before a certain time. For example, we may know that a certain device failed sometime before 200 but we do not know exactly when.

3.6.2 Parameter Estimation

No matter what type of data we have, an important issue in failure statistics is to estimate the parameters of a given probability distribution thought to be a good model for the failure data at hand. Several parameter estimation methods are available. In this section, we present an overview of three methods, ranging from the relatively simple graphical *probability plotting* method to the involved *least squares* and *maximum likelihood* methods. We assume that we are dealing with complete data.

3.6.2.1 Probability Plotting

The method of probability plotting takes the cumulative distribution function (CDF) and attempts to linearize it by employing a specially constructed paper. Here we will use the two-parameter Weibull distribution to illustrate the method. Recall from Section 3.5.4 that the *CDF* of the two-parameter Weibull distribution is given by Equation 3.16, namely

$$F(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^\beta}$$

This function can then be put in the common linear form of $y = a + bx$ as follows: rewriting Equation 3.16 as

$$1 - F(t) = \exp\left(-\left(\frac{t}{\theta}\right)^\beta\right),$$

and taking the logarithm of both sides we obtain

$$\ln(1 - F(t)) = -\left(\frac{t}{\theta}\right)^\beta.$$

Taking the logarithm one more time, we have

$$\ln(-\ln(1-F(t))) = \beta \ln\left(\frac{t}{\theta}\right),$$

or

$$\ln\left(\frac{1}{\ln(1-F(t))}\right) = \beta \ln t - \beta \ln \theta.$$

If we let $y = \ln\left(\frac{1}{\ln(1-F(t))}\right)$ and $x = \ln t$ then the equation can be rewritten as

$$y = \beta x - \beta \ln \theta, \quad (3.20)$$

which is now a linear equation with slope β and intercept $-\beta \ln \theta$.

Weibull graph paper can be constructed by relabeling the paper axes as $x = \ln t$ and $y = \ln\left(\frac{1}{\ln(1-F(t))}\right)$. The values x can be computed easily from the data. However, the computation of y requires the estimation of $F(t)$ from the data, which correspond to the fraction of the population failing prior to each sample value. The most commonly used method of determining this value is by obtaining the median rank for each failure.

The median rank is the value that the true probability of failure, $F(t_j)$, should have at the j th failure out of a sample of N units at a 50% confidence level; which means that this is our best estimate for $F(t_j)$. This estimate is based on a solution of the binomial equation.

The rank can be found for any percentage point, P , greater than zero and less than one, by solving the cumulative binomial equation for Z . The variable Z represents the rank, or $F(t_j)$ estimate, for the j th failure (Johnson, 1951) in the following equation for the cumulative binomial:

$$P = \sum_{n=j}^N \binom{N}{n} Z^n (1-Z)^{N-n} \quad (3.21)$$

where N is the sample size and j the order number of the failure. The median rank is obtained by solving Equation 3.21 for Z at $P = 0.50$:

$$0.5 = \sum_{n=j}^N \binom{N}{n} Z^n (1-Z)^{N-n} \quad (3.22)$$

Solving Equation 3.22 for Z requires the use of numerical methods. A quick and less accurate approximation of the median ranks, known as *Benard's approximation*, is given by

$$\text{Median Rank} = \frac{j - 0.3}{N + 0.4} \quad (3.23)$$

Note that the only information needed to compute the median rank and consequently have an approximation of the cumulative distribution is the order of the failure in the sample.

Once the pairs $(t_j, F(t_j))$ are available, they are plotted on Weibull paper as explained earlier. The parameter β of the Weibull distribution is obtained from the slope of the straight line fitted to the plotted points. As to the estimate of the scale parameter θ , it can be obtained in a simple way as follows.

Let us set $t = \theta$ in the CDF equation. Then we have

$$F(\theta) = 1 - e^{-\left(\frac{\theta}{\theta}\right)^\beta} = 1 - e^{-1} = 0.632$$

Therefore, the value of the parameter θ is the value of t on the x-axis that corresponds to the value of 63.2% on the y-axis.

Example 3.16: The time to failure of six identical components is given in Table 3.3.

Table 3.3. Data for Example 3.16

Failure	1	2	3	4	5	6
Time to Failure	46	95	112	198	325	665

Using the graphical method, estimate the Weibull shape parameter, β and the characteristic life θ .

In order to use Weibull paper to estimate the parameter we need the failure times and the median ranks. These are summarized in Table 3.4

Table 3.4. Median ranks (calculated using Equation 3.23) for the example data

i	t_i	Median rank
1	46	0.109375
2	95	0.265625
3	112	0.421875
4	198	0.578125
5	325	0.734375
6	665	0.890625

Figure 3.12 shows the graphing of the data in Table 3.4 on Weibull paper. From the graph $\beta = 1.2$ and $\theta = 270$.

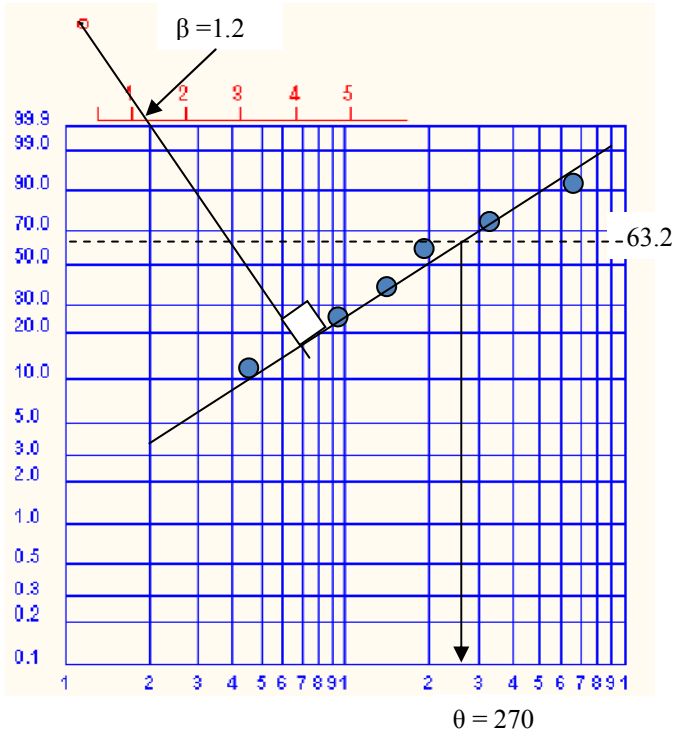


Figure 3.12. Graphical estimation of Weibull parameters for Example 3.16

Probability plotting requires a lot of effort and is not always consistent, as it is not easy to draw the line that best fits a set of points. It was used before the widespread use of computers that could easily perform the calculations for more sophisticated methods, such as the least squares and maximum likelihood methods, which are discussed next.

3.6.2.2 Least Squares Method

The method of least squares requires that a straight line be fitted to a set of data points, such that the sum of the squares of the distance of the points to the fitted line is minimized.

Assume that a set of data pairs $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, were obtained and plotted, and that the x -values are known exactly. Then, according to the *least squares principle*, which minimizes the vertical distance between the data points and the straight line fitted to the data, the best fitting straight line to these data is the straight line $y = \hat{a} + \hat{b}x$, where $\hat{a}$ is an estimate of a and $\hat{b}$ is an estimate of b . These estimators are obtained by minimizing the least squares function

$$L(a, b) = \sum_{i=1}^N (a + bx_i - y_i)^2 \text{ and are given by}$$

$$\hat{b} = \frac{\sum_{i=1}^N x_i y_i - \frac{\sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N}}{\sum_{i=1}^N x_i^2 - \frac{\left(\sum_{i=1}^N x_i\right)^2}{N}} \quad (3.24)$$

and

$$\hat{a} = \frac{\sum_{i=1}^N y_i}{N} - \hat{b} \frac{\sum_{i=1}^N x_i}{N} = \bar{y} - \hat{b} \bar{x} \quad (3.25)$$

Note that the least squares estimation method is best used with data sets containing complete data with no censored or interval data.

Example 3.17: In this example we illustrate how least squares method is used to estimate the parameter of the exponential distribution. Assume that we have a complete data set of n failures, $t_1, t_2, \dots, t_n$.

For the exponential distribution, the equations for y_i and x_i are

$$y_i = \ln[1 - F(t_i)]$$

and

$$x_i = t_i$$

and the $F(T_i)$ is estimated from the median ranks.

Using Equations 3.24 and 3.25, we obtain

$$\hat{a} = 0$$

and

$$\hat{b} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$$

Example 3.18: Consider the data in Example 3.16. Let us use the least squares method to estimate the parameters of the Weibull distribution.

The computation of the y_i s and x_i s are summarized in Table 3.5.

Table 3.5. Least squares computations

		x_i		y_i	$x_i y_i$	x_i^2
i	t_i	$\ln(t_i)$	Median rank	$\ln\left(\ln\left(\frac{1}{1-F(t_i)}\right)\right)$		
1	46	3.8286	0.1094	-2.1556	-8.2531	14.6585
2	95	4.5539	0.2656	-1.1753	-5.3520	20.7378
3	112	4.7185	0.4219	-0.6015	-2.8384	22.2642
4	198	5.2883	0.5781	-0.1473	-0.7789	27.9658
5	325	5.7838	0.7344	0.28192	1.6306	33.4526
6	665	6.4998	0.8906	0.79434	5.1630	42.2472
Sums		30.6729		-3.0035	-10.429	940.827

Using Equations 3.24 and 3.25:

$$\hat{b} = 1.09$$

$$\hat{\beta} = \hat{b} = 1.09$$

$$\hat{a} = -6.069$$

$$\hat{a} = -\hat{\beta} \ln \hat{\theta} = -6.069$$

Hence $\hat{\theta} = 263$.

3.6.2.3 Maximum Likelihood Method

For a given distribution, the maximum likelihood method tries to obtain the most likely values of the parameters that will best describe the data.

If X is a continuous random variable with probability density function $f(x, \theta_1, \theta_2, \dots, \theta_k)$, where $\theta_1, \theta_2, \dots, \theta_k$ are the parameters of the distribution that need to be estimated and $x_1, x_2, \dots, x_N$ are N independent observations which corresponds for example to failure times, then the likelihood function is given by

$$L(\theta_1, \theta_2, \dots, \theta_k \mid x_1, x_2, \dots, x_N) = \prod_{i=1}^N f(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

The logarithmic likelihood function is given by

$$\ln L(\theta_1, \theta_2, \dots, \theta_k \mid x_1, x_2, \dots, x_N) = \ln \prod_{i=1}^N f(x_i; \theta_1, \theta_2, \dots, \theta_k) = \sum_{i=1}^N \ln f(x_i; \theta_1, \theta_2, \dots, \theta_k).$$

The maximum likelihood estimators of $\theta_1, \theta_2, \dots, \theta_k$ are obtained by maximizing L or $\ln L$, which is much easier. Therefore the MLE estimators of are solutions to the simultaneous equations

$$\frac{\partial \ln L}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, k.$$

Example 3.19: Let X be exponentially distributed with parameter θ . Assume that we have a complete data set of n failures, $t_1, t_2, \dots, t_n$. The likelihood function is given by

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{t_i}{\theta}} = \frac{1}{\theta^n} \exp \left[-\sum_{i=1}^n t_i / \theta \right].$$

The log likelihood function is

$$\ln L(\theta) = n \ln \frac{1}{\theta} - \frac{1}{\theta} \sum_{i=1}^n t_i.$$

Taking derivative with respect to the parameter θ and solving for θ , we obtain the MLE estimator

$$\hat{\theta} = \frac{\sum_{i=1}^n t_i}{n}$$

which is the sample mean.

3.6.2.4 Analysis of Suspended Data

As discussed earlier, items are sometimes taken off test for reasons other than failure. For example, we may intentionally place more items than we intend to fail to reduce testing time. However, all available data should be considered in the analysis of times-to-failure data. To accommodate suspensions in the data, we assign an average order number to each failure time. The analysis of suspended data is illustrated using Example 3.20 (Kapur and Lamberson, 1977).

Example 3.20: Assume that four items are placed on test with results shown in Table 3.6

Table 3.6. Example 3.20 data

Failure or Suspension	Symbol	Hours on test
Failure	F ₁	84
Suspension	S ₁	91
Failure	F ₂	122
Failure	F ₃	274

If the suspended item had continued to failure, then we will have three possible outcomes for the order of the failures, depending on when the suspended item would have failed as shown in Table 3.7.

Table 3.7. Possible failure time of suspended item

Outcome 1	Outcome 2	Outcome 3
F_1	F_1	F_1
$S_1 \rightarrow F$	F_2	F_2
F_2	$S_1 \rightarrow F$	F_3
F_3	F_3	$S_1 \rightarrow F$

Notice that:

- The first observed failure time is not affected by the suspension and its order number will always be $j = 1$;
- For the second failure time, it can have either an order number $j = 2$ (two ways) or an order number $j = 3$ (one way); and
- Third failure time can have either order number $j = 3$ (one way) or order number $j = 4$ (two ways).

Therefore the order number of the first failure time is $j = 1$. An average order number is assigned to the second and third failure times as follows:

The average order number for the second failure time is given by:

$$j = \frac{2 \times 2 + 3 \times 1}{3} = 2.33.$$

The average order number for the third failure time can be obtained in a similar manner. The order number of the three failure times and their median ranks are summarized in Table 3.8.

Table 3.8. Order numbers and adjusted median ranks for Example 3.20

Failure	Hours on test	Order Number	Median Rank
			$\frac{j-0.3}{n+0.4} = \frac{j-0.3}{4+0.4}$
F_1	84	1	0.159
F_2	122	2.33	0.461
F_3	274	3.67	0.766

Finding all possible sequences for a mixture of several failures and suspensions in order to calculate the average order number for each failure would be a very time consuming task. Fortunately, there is simple formula for calculating order numbers (Johnson, 1964). The method uses Equation 3.26 for computing increments.

$$Inc = \frac{(n+1) - (\text{previous order number})}{1 + (\text{number of items following suspended set})} = 2.33 \quad (3.26)$$

The method is best illustrated though Example 3.21.

Example 3.21: Consider the data in Table 3.9.

Table 3.9. Data for Example 3.21

Hours on test	Failure or suspension
500	F ₁
620	F ₂
780	S ₁
830	S ₂
850	F ₃
970	F ₄
990	S ₃
1150	F ₅

The first two failure times will have order numbers 1 and 2, respectively. However for the third failure time, an increment must be calculated to account for preceding suspensions. Using Equation 3.26, the increment is

$$Inc = \frac{(8+1)-(2)}{1+(4)} = 1.40$$

To obtain the order number of the third failure, we add this increment to the previous order number which is 2. Therefore the order number of the third failure is $2 + 1.40 = 3.40$. We continue with the same increment until the next suspension set is encountered. Hence the order of the fourth failure is obtained by adding the same increment, 1.80, to the order number of the third failure. Hence the order number of the fourth failure is $3.80 + 1.40 = 4.80$. To compute the order number of the fifth failure, we need to compute a new increment because of the third suspension. The new increment is:

$$Inc = \frac{(8+1)-(4.80)}{1+(1)} = 2.10$$

Therefore the order number of the fifth failure is $4.80 + 2.10 = 6.90$. The order numbers and median ranks for the five failure times are summarized in Table 3.10.

Table 3.10. Order numbers and adjusted median ranks for Example 3.21

Failure	Hours on test	Order Number	Median Rank
			$\frac{j-0.3}{n+0.4} = \frac{j-0.3}{4+0.4}$
F ₁	500	1	0.083
F ₂	620	2	0.202
F ₃	274	3.40	0.369
F ₄	850	4.80	0.536
F ₅	1150	6.90	0.786

This data can be plotted or used in least squares method to estimate distribution parameters.

References

<http://www.weibull.com/>

Johnson, LG, (1951) The Median Ranks of Sample Values in their Population With an Application to Certain Fatigue Studies. Ind Math, Vol. 2, 1951.

Johnson, LG, (1964) Theory and Technique of Variation Research. Elsevier Publishing, New York, NY.

Kapur KC and Lamberson LR (1977) Reliability in Engineering Design. John Wiley, New York.

Montgomery DC and Runger CR (1999) Applied Statistics and Probability for Engineers, 2nd Ed., John Wiley, New York.

Ross SM (1980) Introduction to Probability Models. Academic Press, New York.

Maintenance Control Systems

Maintenance Control

Salih O. Duffuaa and Ahmed E. Haroun

5.1 Introduction

A maintenance system can be viewed as a simple input/output system. The inputs to the system are manpower, failed equipment, material and spare parts, tools, information, policies and procedures, and spares. The output is equipment that is up, reliable and well configured to achieve the planned operation of the plant. The system has a set of activities that make it functional. The activities include planning, scheduling, execution and control. The control is achieved in reference to the objectives of the maintenance system. The objectives are usually aligned with the organization objectives and include equipment availability, costs and quality. The feedback and control is an important function in this system that can be used to improve the system performance. A typical maintenance system with key processes and control function is shown in Figure 5.1. The figure exemplifies the role and the need for effective feedback and control.

An effective maintenance control system improves equipment reliability and assists in the optimal utilization of resources. Maintenance control refers to the set of activities, tools and procedures utilized to coordinate and allocate maintenance resources to achieve the objectives of the maintenance system that are necessary for the following:

1. Work control;
2. Quality and process control;
3. Cost control; and
4. An effective reporting and feedback system.

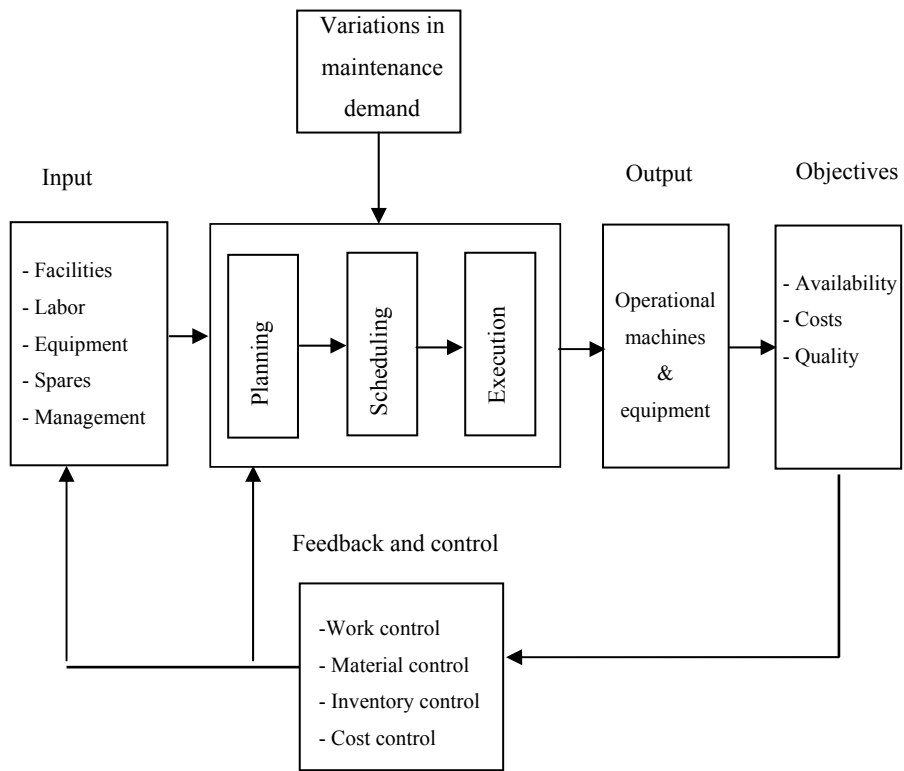


Figure 5.1. Maintenance system and process control

An essential part of the maintenance control is the work order system that is used for planning, executing and controlling maintenance work. The work order system consists of necessary documents and well defined flow process of the work order. The documents provide the means for planning and collecting the necessary information for monitoring and reporting maintenance work. Maintenance control has received a considerable interest in the literature. Duffuaa *et al.* (1998), Neibel (1984) and Kelly (1984) each devoted a chapter in their books on maintenance control. Al-Sultan and Duffuaa (1995) advocated the use of mathematical programming to accomplish effective maintenance control. Gits (1994) presented a detailed structure for maintenance control.

This chapter covers the elements and structure of maintenance control. It presents the required functions for effective control. Section 5.2 describes the maintenance control as a management function followed by the steps of the maintenance control process in Section 5.3. Section 5.4 presents the functional structure of maintenance control followed by the work system in Section 5.5. Section 5.6 outlines some of the necessary tools for developing effective maintenance control and section 5.7 suggests a set of programs that may be employed to improve maintenance control. Section 5.8 provides a brief summary for the chapter.

5.2 The Maintenance Control Function

The maintenance control function can be viewed as an important and integral part of the maintenance management function (MMF). The MMF consists of planning, organizing, leading and controlling maintenance activities (Schermehorn, 2007). The planning function develops objectives and targets to be achieved. In the case of maintenance the targets could be measures regarding availability, quality rates and production. Then management organizes, provides resources and leads to perform tasks and accomplish targets. The implementation of the plans are undertaken to accomplish intended objectives. The fourth function of maintenance management is controlling which concerns monitoring, measuring performance, assessing whether objectives are met and taking necessary corrective actions if needed. Figure 5.2 depicts the management function, its sub-functions and their interactions.



Figure 5.2. Maintenance control as a function of the management process

Maintenance managers must have the capabilities to recognize performance problems and opportunities, make good decisions, and take appropriate action in order to achieve organization success in terms of performance effectiveness and efficiency, and hence of attainment of a high level of productivity.

Maintenance managers and planners maintain active contact with personnel in the course of their work, gather and interpret reports on performance/goals attainment and efficient utilization of the resources (materials, man-hours, and time of job performed) and use information to plan constructive actions in order to control maintenance.

Effective control is important to organizational learning. The follow-up, review, monitoring and streamlining of the practice (corrective actions) makes continuous improvement become a genuine part of organizational culture. It encourages everyone involved in the maintenance process to be responsible for their performance efforts and accomplishments.

5.3 The Control Process

The process of maintenance control involves four steps as shown in Figure 5.3:

1. *Establish objectives and standards*: the control process begins with planning, when performance objectives and standards to be measured are then set. Performance objectives should represent key (essential) results that must be accomplished.
2. *Measure actual performance*: the goal is to accurately measure the performance results (output standards) and/or the performance efforts (input standards). Measurement must be accurate enough to pinpoint significant differences between what is actually obtained and what was originally planned. In maintenance performance measurements, the following indices assist in setting targets and assessing if they are met or not:

(a) Production Indices:

Quality rate (QR) = (Units produced within specifications)/(Total units produced).

Process rate (PR) = (Speed of machine operation)/(Design speed)

Machine Utilization (U) = (Actual production achieved (hrs))/(Total scheduled hours)

Percentage lost production due to causes other than Maintenance = (Lost production hours due to causes other than breakdowns)/(Total lost production hours)

(b) Maintenance Indices:

Overall equipment effectiveness (OEE) = $U \cdot PR \cdot QR$

Percentage of lost production hours due to breakdown = (Lost production hours due to breakdown)/(Total lost production hours)

Mean time between failures (MTBF) = (Number of available operating hours)/(Number of breakdowns)

Mean time of repair (MTR) = (Sum of all repair times)/(Number of of breakdowns)

Machine breakdown severity = (Cost of breakdown repairs)/(Number of of breakdowns)

Percentage of planned maintenance = (Number of maintenance hours worked as planned)/(Total maintenance hours worked)

Maintenance efficiency = (Actual hours worked on aintenance)/(Total available hours for maintenance)

Effective cost of labor/hour = (Total cost of labor (wages and overtime))/(Actual hours worked)

Effective cost of maintenance/man/hour

(1) = (total cost of maintenance)/(total man hours worked)

(2) = (direct cost of breakdown repairs (labor and material)/(total direct cost of all maintenance)

For more on indices refer to Chapter 2 on maintenance productivity in this book.

3. *Compare results with objectives and standards*: this step can be expressed in the Control Equation: *Need for Action = Desired Performance – Actual Performance*. Sometimes managers make a *historical comparison*, using past performance as a basis for evaluating current performance. A *relative comparison* uses the performance achievements of other persons, work units, or organizations as the evaluation benchmarks. In *maintenance comparisons* standards are set scientifically through such methods such as time and motion studies. The preventive maintenance routines, for example, are measured in terms of expected time in every routine performed, based on operating hours, or time interval (see Figure 5.3).
4. *Take corrective action*: the final step in the control process is to take any action necessary to correct problems, discrepancies, or make improvements. Management by exception is the practice of giving attention to situations that show the utmost need for action. It saves valuable time, energy, and other resources by focusing attention on critical and high-priority areas. The maintenance managers should give special attention to two types of exceptions: 1) a *problem situation* in which actual performance is below the standard; and 2) an *opportunity situation* in which actual performance is above the standard. The reason for this is that with the goal of existence, enterprises should look for achieving a high level of productivity.

5.4 Functional Structure of Maintenance Control

In Section 5.2 we viewed the maintenance control as one of the management functions. In this section the functional structure of maintenance control will be described. The structure of maintenance control consists of the following important functions:

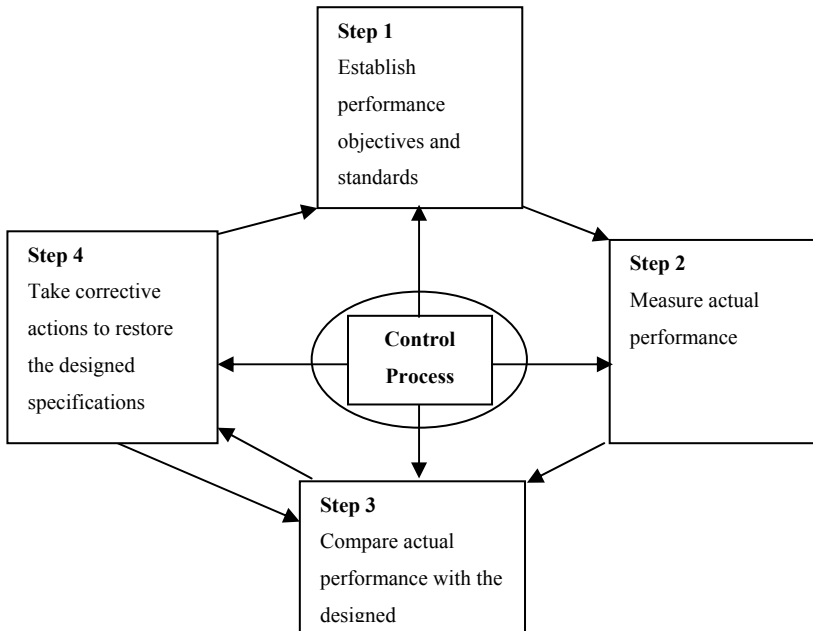


Figure 5.3. Four steps of the maintenance control process

1. *Planning and forecasting the maintenance load:* the planning and forecasting of maintenance load deals with two important aspects of maintenance. The first aspect is the emphasis on planning the maintenance load which is a result of a planned maintenance program. The second aspect deals with forecasting the maintenance load. The functions of planning and forecasting (predicting) maintenance load are prerequisites for effective maintenance control and are dealt with in detail in Chapters 8 and 11 in this book. However, in this chapter it is important to mention that the best way of predicting the maintenance requirements is to have a large portion of the maintenance load planned. This necessitates an effective planned maintenance program that ensures at least 80% of the maintenance load is planned and it's preferable to have 90% of the load planned. Unplanned maintenance work is a major factor in lack of control unlike planned maintenance work that reduces uncertainties in planning the required resources and coordination to accomplish the maintenance work and hence assists in effective maintenance control.
2. *Work order planning and scheduling:* the functions of work order planning and scheduling deal with planning the resources for the required maintenance jobs and allocating the available resources. The resources include manpower, material, spare parts and tools. Usually this requires a job of a planner who is well training in productivity methods, time standards, materials, computers and has good communications skills. The scheduling deals with allocation of the available resources at specified points in time. The work order planning requires the existence of a well designed work order system (See 2.6 below).

Then scheduling the work based on an established priority system. A brief outline of these functions is provided in this chapter since they are presented in detail in a separate chapter of this book.

3. *Work order execution and performance evaluation:* the work order execution and monitoring functions deal with processing the work orders and monitoring the progress of the work through the work cycle. In this function data is collected to assess the quality of work and the utilization of the resources.
4. *Feedback and corrective action:* feedback information and corrective action is concerned with the collection of data about the status of the work execution, system availability, work backlog, quality of work performed. Then this information is analyzed and communicated to decision makers in order to take appropriate corrective actions and, thus, to aid in achieving set goals and objectives.

5.5 Work Order System

The work order system consists of two main parts: (1) the documents required to facilitate work planning, execution and control; and (2) the work order flow process.

5.5.1 Basic Documentation for Work Order System

The necessary documents required for the work order system include the work order, materials and tools requisition forms, job card, maintenance schedule, maintenance program, plant inventory and equipment history files. Descriptions and examples are provided below.

5.5.1.1 The Work Order (W/O)

The work order is the basic document (form) for planning and control. It is necessary to ensure that any request, failure and remedy are recorded for further use (Figure 5.4 is an example of a typical W/O). In industry, W/Os may be referred to by different names such as work request, work requisition, request for service, *etc.* The W/O can be initiated by any persons in the organization and must be screened by the maintenance planner or coordinator. Detailed written instructions for any work or activities (job) to be carried out, in any component or part of a plant/equipment/machinery, must be clearly shown in the W/O. So, the work order is used for the following:

WORK ORDER										
Work order No. Requesting Dept Date Time Plant Description Plant Register Card #					Plant Location: Department Unit Cost Center..... Shift: Morning ☐ Afternoon ☐ Night ☐					
DEFECT/WORK REQUIRED										
PRIORITY		Emergency ☐ Urgent ☐ Normal ☐ SCHEDULED: Preventive ☐ Predictive ☐								
CAUSE		Wear & Tear ☐ Accident/Misuse/Neglect ☐ Component Failure n N/A ☐								
DETAILS OF CAUSE:										
Tradesman (Labor)					Materials					
Time/cost					Parts		Cost			
Trade Code	Est Time	Actual Time	Trade hourly cost	Total cost	Description	Part No.	# of Units	Unit Price	Total Cost	
Total Repair Time		Hours		Minutes	Total Materials Costs					
					Total Labor Costs					
					TOTAL COSTS					
Technician Signature					Date Completed					
Job Approval.....					Date Approved.....					

Figure 5.4. Work order form

1. Detailing the required resources for the job including the assignment of skilled and competent personnel for undertaking the maintenance tasks;
2. Ensuring appropriate and the best methods and procedures utilization including safety procedures;
3. Execution, maintaining, monitoring and controlling the maintenance activities and tasks; and
4. Providing the right data and information from the work order for analysis and continuous improvements.

The processing of the work order system is the responsibility of the persons in charge of planning and scheduling. Therefore, the work order is designed to include all necessary information needed to facilitate effective planning and scheduling and control. Information needed for planning and scheduling include the following:

- Inventory number, unit description and site;
- Person or department requesting the work and date work required;
- Work description and time standards;
- Job specification, code number and priority;
- Crafts required;
- Special tools;
- Safety procedures; and
- Technical information (drawings and manuals).

Information needed for control include:

- Actual time taken;
- Cost codes for crafts;
- Down time or time work finished; and
- Cause and consequences of failure.

5.5.1.2 Materials and Tools Requisition: Figures 5.5 and 5.6

The W/O should be supplemented by two requisition forms; one for materials (Figure 5.5) and the other for tools (Figure 5.6). Those forms are necessary to ensure that materials and tools are ready before the job is started.

These two forms are also useful for providing information to facilitate smooth and timely planning and control. Such information includes:

- Inventory number, unit description and site;
- Work description and time standards;
- Job specification and code number;
- Spare parts and material required;
- Special tools required;
- Stock control;
- Stores code and units price; and
- Time required for tools use.

5.5.1.3 Job Card

The job card (Figure 5.7) describes the maintenance plan for specific equipment. It carries time taken for repair, inspection or preventive maintenance.

MATERIALS REQUISITION						
Work order No. Requesting Dept Date Time Plant Description Plant Register Card #			Plant Location: Department Unit Cost Center..... Shift: Morning ☐ Afternoon ☐ Night ☐			
Materials Requirement					Price	
Store Code	Group	Part No.	Materials List Description	Qty Req'd	Unit	Total
Storekeeper Initials:		Delivered To:		Stock Control Entered By:		Received By:
Date Delivered:			Date Received:			

Figure 5.5. Materials requisition form

5.5.1.4 Plant Inventory

Lists all plant items and allocates each item an individual code number. The plant inventory should be supplemented by a front page, containing the technical details about the plant/equipment/machinery, and could be called a Plant Register or Card.

5.5.1.5 Maintenance Schedule

A comprehensive list of maintenance and its incidence (frequency of occurrence) over the life cycle of the assets is a general guideline to assist in developing routine maintenance. For example, in the case of motor vehicles, it assists with a vehicle's routine maintenance at a set odometer reading or time schedule, depending the use and driving habits. A comprehensive list for a university includes all assets which require up-keeping, *i.e.*, buildings, transport fleet, air conditioning systems, audio-visuals, stand-by generator set, *etc.*, so as to determine all required activities for the whole life cycle of the different physical assets. Based on the schedule, managers set the appropriate maintenance organization, workforce, out-sourcing policies, and periodic maintenance programs. See Haroun and Ogbugo (1981).

TOOLS REQUISITION			
Work order No. Requesting Dept DateTime Plant Description Plant Register Card #		Plant Location: Department Unit Cost Center..... Shift: Morning ☐ Afternoon ☐ Night ☐	
Tools Requisition			
Store Code	Tools List Description	Job Description	Time Required
Storekeeper Initials:	Delivered To:	Received By:	
Date and Time Received:		Date and Time Returned:	

Figure 5.6. Tools requisition

5.5.1.6 Maintenance Program

This is a plan allocating specific maintenance to a specific time period, often in chart form.

5.5.1.7 Plant History (Record)

Contains information about all work done on plant items including equipment history files. The history file includes work performed, down time and causes of failure.

5.5.2 Work Order System Flow

The work order system flow refers to the dispatching procedures and the order in which the job is processed from its initiation till its completion (Figure 5.8). In this subsection we focus on the work order flow. The following are the sequential steps for the W/O processing:

1. Upon receipt of the work request by the planner (it can be initiated via telephone, computer terminal, or in hard copy) it is screened and checked to determine whether it is a planned maintenance (*i.e.*, preventive or predictive) or an occurrence of a failure.
2. If the job is an EMERGENCY case, then a maintenance crew is dispatched IMMEDIATELY and the W/O follows later.
3. Otherwise a work order is planned and completed, showing the needed information for planning, execution and control *i.e.*, check equipment history, job card, fill-in requisitions for materials and tools, plan manpower, *etc.* Usually three to four copies are directed to the planner, foreman, accountant, and supervisor. This is done online in typical contemporary practice.
4. The foreman of the appropriate unit may give a hard copy to the craftsmen assigned to the job, or the W/O can be accessed directly by crafts through Enterprise Resource Planning (ERP) or Computerized Maintenance Management System (CMMS) equipment. The craftsman completes the job and fills information on the W/O.
5. The foreman checks the quality of work and verifies information and approves or complete his copy on the system (if the system is manual, he puts the verified information on the relevant copies and then forwards them to the maintenance control).
6. Accounting completes costs information on his copy/system.
7. The system extracts data and puts the data in the equipment history file for periodic analysis to control and improve maintenance strategies and policies.
8. The planner verifies the job is completed and all required information extracted and then closes the W/O.

The above steps could be handled manually or automated. Figure 5.8 displays a flow chart showing these steps. If an automated system is used these copies can be stored as copies in the system and circulated online *via* a local area network.

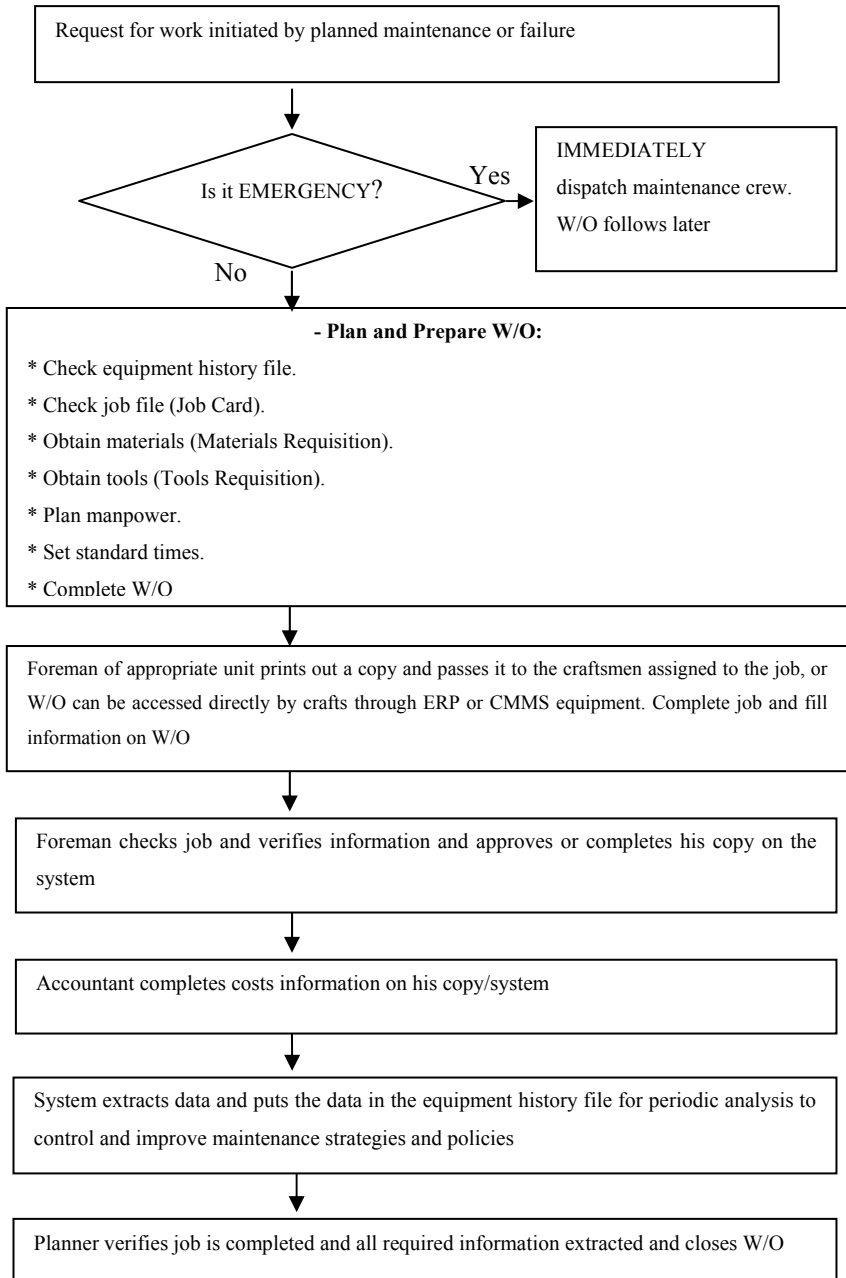


Figure 5.8. Work order (W/O) flow

5.6 Tools Necessary for Effective Maintenance Control System

To achieve reliable maintenance plans and control procedures, the following management techniques and tools should be used:

1. Statistical process control tools (SPC tools): the SPC tools assist in identifying major causes of failures, process capabilities and stability and examine machines and gauges capabilities. The tools include Pareto chart, fishbone diagram (cause and effect diagram) and control charts.
2. Network analysis: network analysis has been successfully used in the power supply and petrochemical industries to reduce plant stoppage by modeling large maintenance jobs, overhauls and plant shutdowns as a network model in order to minimize job completion time and shutdown periods using critical path analysis.
3. Failure mode and effect analysis (FEMA): a failure mode and effects analysis (FMEA) is a procedure for analysis of potential failure modes within a system for the classification by severity or determination of the failure's effect upon the system. It is widely used in the manufacturing industries in various phases of the product life cycle. Failure causes are any errors or defects in process, design, or item especially ones that affect the customer, and can be potential or actual. Effects analysis refers to studying the consequences of those failures; or, applying such analysis in failure's risk assessment for systematically identifying potential failures in a system or a process.
4. FIMS (Functionally Identified Maintenance system): FIMS is a diagnostic technique that represents equipment or a system in a hierarchical logical sequence. In the hierarchical representation each level is a functional and logical development of the preceding one. The purpose of FIMS is to identify the failure location in an easily and timely manner. It has been applied successfully in complex systems such as refineries, airplanes and locomotives.
5. Work measurement: work measurement is one of the elements of work study. It is a technique to develop time standards of jobs while considering ratings of workers and allowances for personal needs, fatigue and other contingencies. Time standards are essential for accurate scheduling, control and incentive schemes.
6. Stock control: effective policies for spare parts and materials ordering play a critical role in reducing down time. Planned maintenance programs facilitate the ordering of spare parts and consumables. The application of economic re-order quantities based on material usage data not only reduces the total inventory cost, but plant down-time and maintenance labor costs.
7. Budget: budgeting is essential for cost control. It forms a basis for the judgment of actual performance, and through cost control it shows if remedial measures are necessary. The real costs of maintenance are not easily assessed. Its true cost should be segregated from those of the

indirect activities of the department, if satisfactory control and accountability are to be established.

8. Life cycle costing (LCC): life cycle cost is the total cost of ownership of machinery and equipment, including its cost of acquisition, operation, maintenance, conversion, and/or decommission (SAE 1999; Barringer, 2003). The objective of LCC analysis is to choose the most cost appropriate approach from a series of alternatives to achieve the lowest long-term cost over the life of an asset. Usually the cost of operation, maintenance, and disposal costs are the major component in the LCC. The use of LCC may help in reduction in the costs of maintenance and operation since these costs are the dominant ones in machinery LCC.
9. Computerized maintenance management systems (CMMS): CMMS enables maintenance managers and supervisors to access information about equipment, manpower and maintenance policies. This information assists in improving maintenance effectiveness and control.

The above tools and techniques are aimed at controlling and improving the following:

5.6.1 Work Control

Work control deals with monitoring the work status and the accomplished work to investigate if the work is done according to standards (quality and time). To achieve this type of control it is assumed that the maintenance control system includes standards that are assigned in advance of performing actual maintenance work. A set of reports are generated in this category of control. These include a report showing performance according to standard by the crafts utilized for the job and their productivity. In this report, it is a good practice to categorize the maintenance work whether it is performed in regular in house, over time or outsourced. Other reports that are useful for work control are backlog, percentage of emergency maintenance to planned maintenance, and percentage of repair jobs that originated as a result of PM inspection.

The backlog report is very essential for work control. It is good practice to maintain a weekly backlog report by craft and to indicate the backlog cause. It is also good practice to have a healthy backlog. The size of a healthy backlog ranges between 2–4 weeks. An excessive or too little backlog necessitates a corrective action. In case a down trend in the backlog is identified, *i.e.*, it keeps decreasing, one of the following actions may be necessary as described in Duffuaa *et al.* (1998):

1. Reduce contract maintenance;
2. Consider transfer between departments or crafts; and
3. Down size the maintenance force.

If there is an increasing trend in the backlog, a corrective action is needed which may include one of the following:

1. Increase contract maintenance;
2. Transfer between departments or crafts;
3. Schedule cost effective overtime; and
4. Increase maintenance workforce.

5.6.2 Cost Control

The maintenance cost consists of the following categories:

1. Direct maintenance cost (costs of labor material, spares, material, and equipment);
2. Operation shutdown cost due to failure;
3. Cost of quality due to product being out of specification, as a result of machines' incapability;
4. Redundancy cost due to equipment backups;
5. Equipment deterioration cost due to lack of proper maintenance; and
6. Cost of over maintaining.

Almost all information about cost is available on the worker order. A summary of maintenance costs by work category must be issued monthly. This is utilized to control maintenance costs and develop costs of manufactured products.

The areas where cost reduction programs can be launched to reduce maintenance cost are:

1. Considering the use of alternative spare parts and materials;
2. Modifying inspection procedures; and
3. Revising maintenance policies and procedures, particularly making adjustments in size of crew and methods.

5.6.3 Quality Control

Maintenance has a direct link to the quality of products as demonstrated by Ben Daya and Duffuaa (1995). Well maintained equipment produces less scrap and improves process capability.

A monthly report on the percentage of repeat jobs and product rejects may help identify which machine requires an investigation to determine causes of quality problems. Once the machines are investigated, a corrective course of action will be taken to remedy problems.

5.6.4 Plant Condition Control

Plant condition control requires an effective system for recording failures and repairs for critical and major equipment in the plant. This information is usually obtained from the work order and equipment history file. The records in the equipment history file include the time of failure, the nature of failure, and the repairs undertaken, total downtime, and machines and spares used.

A monthly maintenance report should include measures on plant reliability. Such measures include mean time between failures, overall equipment effectiveness (OEE) and downtime of critical and major equipment. If a down trend is observed on OEE, downtime and readiness is low, and a corrective action must be taken to minimize the occurrence of failure. The corrective action may require establishment of a reliability improvement program or a planned maintenance program, or both.

5.7 Effective Programs for Improving Maintenance Control

In this section, four engineered maintenance programs are briefly outlined. These programs offer sound courses of action that can be adopted to enhance maintenance control. The objective of these programs is to improve plant availability, reduce cost, and improve OEE and product quality. These programs are listed below:

1. Emergency maintenance;
2. Reliability improvement;
3. Total productive maintenance; and
4. Computerized maintenance management.

5.7.1 Emergency Maintenance

Emergency maintenance refers to any job that should be attended to immediately. Emergency maintenance, by its nature, allows very little lead time for planning. The amount of emergency maintenance must be minimized and it should not exceed 10% of the total maintenance work. The maintenance department must have a clear policy for handling emergency maintenance. One of the following offers an approach to handling emergency maintenance:

1. Preempt the regular schedule and perform the emergency maintenance, then pick up the backlog with overtime, temporary workers or contract maintenance; and
2. Assign dedicated crafts for emergency maintenance based on the estimated emergency maintenance load. It is an accepted practice in industry to allow 10–15% of load capacity for emergency work.

The first approach is expected to result in increased workforce utilization; however, the second approach offers the ability to respond quickly as needed.

5.7.2 Reliability Improvement

A reliability engineering program offers a sound alternative for improving the maintenance function. It can be used as an option to improve maintenance performance. Critical and major equipment history files must be maintained and estimates for mean time between failure (MTBF) must be calculated. The

frequency of emergency maintenance is a function of the failure rate of this equipment. It can be estimated for a period of operations lasting n hours, there will be n/MTBF emergency maintenance actions. The longer the MTBF the lower the number of emergency maintenance incidents.

Reliability centered maintenance (RCM) can be utilized to enhance maintenance polices and improve equipment reliability. In RCM, the maintenance program is developed on the basis of the concept of restoring equipment function rather than bringing the equipment to an ideal condition. RCM has been applied successfully in the commercial airline industry, nuclear reactors, and other power plants.

5.7.3 Total Productive Maintenance

Total Productive Maintenance (TPM) is an approach to maintenance developed in Japan that brings the tools of total quality management (TQM) to maintenance. The aim of TPM is to reduce six categories of equipment losses to improve overall equipment effectiveness (OEE). The six major causes of equipment losses, according to Nakajima (1988) are:

1. Failure;
2. Set-up and adjustments;
3. Idling and minor stoppage;
4. Reduced speed;
5. Process defects; and
6. Reduced yield.

TPM empowers operators and uses multi-skilled crafts to minimize response time and perform productive maintenance. The implementation is expected to assist in improving maintenance effectiveness and control.

5.7.4 Computerized Maintenance Management and Information Technology

High technology production units (machinery/equipment) require high technology maintenance and control systems, so maintenance systems must move in new directions if manufacturers/service enterprises hope to keep that expensive production/service equipment up and running.

Information technology hardware and software enable maintenance management to automate and process activities in a speedy manner. Above all enable maintenance managers to retrieve and process information that can be used for effective maintenance planning and control. Every company must use a basic computerized maintenance system. It is most effective to integrate such systems with organization enterprise resource planning (ERP). Many of the existing ERPs have maintenance modules.

Finally, any system that is installed should serve the maintenance personnel rather than forcing these people to serve the system, so all such systems will require extensive personnel training.

5.8 Summary

Maintenance control systems play a key role in having an effective maintenance program. In this chapter, maintenance control is viewed as an integral part of maintenance management and the steps of effective process control are described. Then the functional structure of maintenance control is explained in detail. The structure consists of maintenance work load forecasting, effective planning and scheduling, work order execution and performance evaluation, and feedback and corrective action. The steps for implementing maintenance control are:

1. Train the maintenance personnel on the concepts and techniques of maintenance control.
2. Develop clear work plans including objectives and targets to be achieved on daily and weekly bases. Also establish standards and measures to assess the progress towards achieving plans and targets.
3. Coordinate, plan and process work orders.
4. Monitor and collect information from work orders and history files and compile reports on efficiency, availability and quality.
5. Examine the deviation from established objectives and targets.
6. If a deviation exists, take corrective action or otherwise revise and set higher targets.

The six programs described in Section 5.6 offer ways and means for improving the effectiveness, efficiency and quality of maintenance, and satisfaction of employees.

Acknowledgement

The authors would to acknowledge the support provided by King Fahd University of Petroleum and Minerals for conducting this work. The effort of Mr. Amr A. E. Haroun for drafting the figures of this chapter is deeply appreciated.

References

- Al-Sultan KS, Duffuaa SO (1995) Maintenance Control Via Mathematical Programming. *J of Qual in Maint Eng* 1:3–20
- Barringer PH (2003) Summary of Life Cycle Costing. *Proc. International Conference of Maintenance Societies, Australia*: 1–10
- Ben Daya M, Duffuaa SO, (1995) Maintenance and Quality: The Missing Link. *J of Qual in Maint Eng* 1: 20–26
- Duffuaa SO, Raouf, A, Campbell JD (1998) *Planning and Control of Maintenance Systems: Modeling and Analysis*. Wiley, New York.
- Gits CW, (1994) Structuring Maintenance Control. *int J of operations and prod manag Special Issue on Maint Manag* 14: 5–17.
- Haroun AE, Ogbugo MA (1981) *Development of Maintenance Policies and Procedures for Crosfield Electronics at Milton Keynes*. CIT, Cranfield, England
- Kelly A (1984) *Maintenance Planning and Control*, Butterworth, London
- Nakajima S (1988) *Introduction to TPM*. Productivity Press, Portland, Oregon

- Neibel BW (1984) Engineering Maintenance Management, 2nd edn. Marcel and Dekker, N.Y.
- Schermerhorn JR (2007) Management, 9th edn. John Wiley, New York
- Society of Automotive Engineers (1999) Reliability and Maintainability: Guideline for Manufacturing Machinery and Equipment M-110.2. Warrendale, PA

Guidelines for Budgeting and Costing Planned Maintenance Services

Mohamed Ali Mirghani

6.1 Introduction

Effective planned maintenance management enables an organization to gain uptime – the capacity to produce and provide goods and services to customers' satisfaction, consistently. This becomes quite critical in capital intensive organizations because of the heavy investment in capital assets needed for serving customers.

Planned (preventive) maintenance involves the repair, replacement, and maintenance of equipment in order to avoid unexpected failure during use. The primary objective of planned maintenance is the minimization of total cost of inspection and repair, and equipment downtime (measured in lost production capacity or reduced product quality). It provides a critical service function without which major business interruptions could take place. It is one of the two major components of maintenance load. The other component is unplanned (unexpected) maintenance. Planned maintenance could be time or use-based or could be condition-based.

An organization's maintenance strategy has to be in line with its business strategy. Quality and the drive for continuous improvement in world-class organizations are changing the philosophy and attitude toward maintenance. Total productive maintenance (TPM) is one of the outcomes of productivity improvement targets aimed at increasing uptime, improving quality, and achieving cost efficiency. The journey to world-class level of excellence indicates that maintenance managers must take a leadership role in improving the maintenance function.

Several capital intensive manufacturing and service organizations started looking at maintenance as a source of revenue by placing their engineering and maintenance units into an arms-length business relationship with operating departments in an effort to make them more competitive. The change is made with the objective of giving these organizations the option of outsourcing their in-house

maintenance services. Such internal competitive pressure is the best incentive for these units to become more competitive and profitable.

TPM and total quality management (TQM) are processes geared towards making a company more competitive. They both involve interrelationships among all organizational functions for continuous improvement purposes. The real power of both TPM and TQM is to use the knowledge base and experience of all workers to generate ideas and contribute to the goals and objectives of the organization.

A prerequisite to improving maintenance and equipment efficiency is human resources development through training. A maintenance improvement program should require all involved employees to participate in training courses that focus on what good maintenance and operation practices include and the rationale for what needs to be improved in the organization.

The problem of overstocking spare parts needed for maintenance could be avoided by applying the same principles of just-in-time (JIT) systems used in manufacturing. In the case of expensive spare parts, it is important to have the least amount of spares that are consistent with management's specification of the likelihood of equipment availability for use when needed.

Properly designed and implemented budgeting and costing systems have a major role to play in improving the effectiveness and efficiency of the maintenance function.

What follows is an overview and guidelines of business budgeting and costing systems of planned maintenance services as a valuable contributor to the organization's overall cost efficiency and profitability.

6.2 An Overview of Budgeting and Costing Systems

6.2.1 Budgeting Systems

A budget is a quantitative expression of a plan and is an aid to the coordination and implementation of this plan. In addition to instilling the discipline of systematic planning into the organization, the budgeting system provides a two-way channel of communication for the various echelons of the organizational hierarchy. This two-way communication capability (top-down and bottom-up) is directly linked to the iterative nature of the budgetary process through which the technical and financial feasibilities of planned actions are assessed. Furthermore, well-formulated budgets provide a sound basis for evaluating departmental and managerial performance.

A budget should not be perceived by the manager as only a mechanism for securing departmental funding. Such a perception would make the budgetary process a number crunching exercise. A properly functioning budgetary system should help a manager understand that a budget needs to fulfil an organization's mission. This requires department managers responsible for budget development to have a thorough understanding of the organization's mission and their department's role in accomplishing it. This indicates that budget development should be designed so that a department manager's focus is on contributing effectively and efficiently toward carrying out the organization's mission.

In today's competitive environment survival requires businesses to be flexible and innovative mainly through the development of new products and services, while continuously improving productivity and customer care. Building the effects of innovation and continuous productivity improvement into the annual budgets can only be achieved through continuous budgeting that rolls the budget at the end of each quarter for the next four quarters. In this setting, the budget also serves as a vital tool for ensuring that the corporate culture has a unified understanding and commitment to strategic objectives.

Budget performance reports provide valuable feedback for controlling operations and/or for revising plans if circumstances change. These budget performance reports need to be custom-tailored to the appropriate level of responsibility because this has a bearing for their timing, form, content and level of aggregation.

The reports should establish control limits for budget variances so that the manager can focus his attention on significant events. The budgetary control system should keep track, over time, of the behavior of variances that are within the control limits so that upward or downward trends could be discerned and reported in feed-forward reports.

6.2.2 Costing Systems

The word cost means resources consumed or sacrificed to reach an objective. Since the resources at the disposal of an organization are scarce, their efficient utilization is one of the primary objectives of management. Costing refers to the purposeful use of resources. Hence, costing and cost allocation are at the center of managing the scarce resources at the disposal of an organization whether it is for-profit or not. One of the purposes of a costing system is to accumulate cost data and assign these data to cost objects. The assignment of costs to cost objects is accomplished through traceability and/or allocation. Traceability has to be economically feasible, meaning that its cost should be less than the cost of the item(s) to be traced. When a cost item becomes traceable to a cost object, it is classified as part of its direct costs. When a cost is not traceable to a cost object, it is classified as an indirect cost and can be assigned to a cost object through allocation.

Bases for allocation are rank-ordered in terms of rigor as follows:

- Cause and effect;
- Benefits received;
- Ability to bear; and
- Fairness or equity.

Since planned maintenance jobs have different technical specifications resulting in differences in the consumption of maintenance resources, the cost of each job has to be developed separately. The proposed framework in this chapter for costing planned maintenance is based on the techniques of job order costing and activity-based-costing (ABC).

6.3 Proposed Budgetary System

6.3.1 Planned Maintenance Operating Budget

The budgetary system should be driven by the organization's mission statement that provides the framework for strategy analysis (see Figure 6.1). Observe that all arrows are two-directional, indicating that:

1. The various components of the budgetary process are interrelated and affect each other; and
2. The budgetary process is iterative in nature involving top-down and bottom-up communication.

The outcome of strategy analysis is a general statement of objectives that relates to the organization's strategic and long-range plans. This statement is a reflection of top management's expectation of where the organization should be in terms of, for example, market share, competitiveness, profitability, cash flows, and so on, by the end of the budget year. The focus is on key result areas or critical success factors.

The primary objective of planned maintenance is to provide reliability-centered maintenance services. Planned maintenance services should be able to project the number of maintenance jobs for the budget period given the budgeted level of manufacturing, core operations, and marketing activities. Through optimization techniques, planned maintenance services should be able to schedule its work during the budget period to meet reliability factors and improve on them. Effective co-ordination and communication with manufacturing, core operations, and marketing is quite critical at this phase.

The planned maintenance services budget should identify the primary and alternative means for achieving its objectives and the amount of resources needed for each alternative. The resources should include:

- Types and quantities of materials and spare parts;
- Labor skills by headcount;
- Support services;
- Training and manpower development; and
- Maintenance equipment and facilities.

Through the budgetary process, planned maintenance services could justify the acquisition of resources for continuous improvements, improved working conditions, and increasing the level of commitment of the individual worker.

The budgetary process should focus the attention of planned maintenance services management on achieving objectives that are in line with the organization's objectives and mission statement. The budgetary process should synchronize required resources with available resources and identify constraints or bottlenecks that could render the budget technically infeasible. Alleviation of constraints or bottlenecks should be done through optimization techniques. Assuring the technical feasibility of the budget might trigger revisions to the

budgets of manufacturing or core operations, sales, marketing, as well as the general statement of objectives.

The budgetary process should also enable planned maintenance services to coordinate and communicate effectively with materials management to assure the availability of required spare parts/materials in terms of quantity, quality, and timing. This prevents holding excessive inventory and paves the way for a JIT environment.

The planned maintenance services budget should be subjected to rigorous “what if” analysis to recognize explicitly uncertainty and add dynamism to the budgetary process. Through sensitivity analysis, an action plan for coping with changing circumstances of planned maintenance services may be developed. The main advantage of this approach is that the manager experiments with a possible plausible scenarios “on paper” without running the risk of a crisis materializing or an opportunity passing by. This adds significant flexibility to a manager's ability to deal with unexpected situations.

The planned maintenance services budget proposal should be structured to highlight its objectives for the budget period; primary and alternative means of achieving these objectives; resources required for each alternative and related costs. The budget proposal should no longer be an extrapolation of the past, but a management tool with a futuristic orientation.

6.3.2 Financial Budget

On completion of the operating budget, the components of the financial budget will be assembled. The first component is the *capital budget* which justifies the acquisition of capital (long-lived) assets and their relationship to current and future operations. In the case of planned maintenance services, the capital budget should include all capital assets and maintenance facilities to be acquired during the budget period and their impact on current and future operations. Capital budgeting techniques should be used to justify the investment in capital assets.

The second component of the financial budget is the *cash budget* (see Figure 6.1) which synchronizes cash inflows and outflows for the budget period. The major source of cash is operating revenues which appear in the sales budget. The magnitude and timing of cash sales as well as the terms of credit sales and the effectiveness of managing accounts receivable are the major determinants of cash inflows. The payment terms for cash operating expenses and capital assets determine the magnitude and timing of cash outflows. The cash budget is prepared for the year as a whole and should be broken down by month or quarter to ascertain availability of cash for operating and capital expenditures throughout the budget year. The cash budget is a very critical document because it determines the financial feasibility of the operating and capital budgets. A cash surplus (or deficit) could be projected for the budget year and for shorter time periods (months, quarters, *etc.*) so that the treasury department could consider all possible alternatives for handling the projected cash surplus or deficit.

In the case of a persistent cash deficit, its impact on operating and capital expenditures should be assessed by the treasury department. This requires close coordination and communication between the treasury department and

operating/support departments so that necessary revisions to operating and capital budgets and, possibly, top management's general statement of objectives can be made. Available cash should be rationed among operating and support departments by a system of priorities of cash outlays from the standpoint of overall organizational effectiveness. Through this rationing system, each organizational unit would be assigned its "fair share" of the "cash pie" to carry out its activities to contribute positively toward achieving the organization's mission-related objectives.

The implications of the cash rationing system for planned maintenance services is that it will reduce the amount of uncertainty surrounding funding the acquisition of resources needed for planned work during the budget period.

6.3.3 The Budget Cycle

A major prerequisite for an effective budgetary system is a time cycle for effectively carrying out all phases of the budgetary process to produce an approved master budget well before the beginning of the budget year. A budget cycle timetable (calendar) should be prepared to satisfy this prerequisite.

6.3.4 Top Management Support

Top management's support for budgeting as a managerial tool can be manifested in the form of a Corporate Budget Committee (CBC) to be chaired by the Managing Director or General Manager with membership of the managers of all functional areas of the business. CBC's main role is to drive the budget process and set the various criteria to be met at the corporate, divisional, and organizational unit levels. The CBC should operate on a presentation basis where the concerned managers will organize presentations of their respective budget proposals with the assistance of the Corporate Finance Department. CBC will review budget proposals in terms of strategic direction, relevance, and accuracy. The CBC should review the consolidated budget from the overall organization point of view considering all plausible scenarios and ascertain its technical and financial feasibilities. CBC should monitor budgetary performance monthly or quarterly and, accordingly, roll the budget.

The total maintenance (planned and breakdown) budget will be presented to CBC. The planned maintenance budget will be prepared according to the framework presented in this chapter. The breakdown maintenance budget will be based on historical data showing the percentage relationship between breakdown maintenance total actual cost and the total cost of total maintenance.

For example, if that historical percentage relationship is 25%, then the planned maintenance budget is to be divided by 75% to arrive at the total annual maintenance (planned and breakdown) budget. The breakdown maintenance budget could then be derived by deducting the planned maintenance budget amount for the total maintenance budget

	Jan	Feb		Dec	Total
Open balance/facilities:					
Cash sales		-		-	
Debtors collection					
Other sales					
Total cash receipts	-	-		-	-
Total cash available	-				
Operating expenses:					
(see LIST 1 in Figure 6.2)					
Total cash disbursement	-				
1- Net cash receipts/disbursements from operations	-				
(see LIST 2 in Figure 6.2)					
2- Cash from/(used in) financing activities	-	-		-	-
3 - Cash disbursements for investing activities (Capex)					
Net change in cash	-	-		-	
Beginning cash balance	-	-		-	
Projected closing balance	-	-	-	-	

Figure 6.1. Cash budget template

LIST 1	
	Wages, salaries & housing
	Vacation settlement
	Fuel & oil
	Rent
	Electricity
	Spare parts
	Maintenancel
	Maintenance support services
	Other
	L/C's export purchasing (the value of items add to inventory
	CAD export purchasing (the value of items add to inventory
	A/C payables local purchasing (from purchases budget)
	Advertisement
	Interest on loans
LIST 2	
	Bank STRL new
	Bank MTL new
	(Rollover) or repayment of STRLs
	Repayments of MTL
	Refinance repayments
	L/C's and CAD's refinance

Figure 6.2. List 1 and list 2 in Figure 6.1

6.3.5 Budget Performance Reports

Budget performance reports should highlight effectiveness and efficiency of operations. In the case of planned maintenance services the difference between budgeted and actual achievements is a measure of effectiveness (*i.e.*, closeness to accomplishing objectives). Whereas the difference between resources consumed and resources that should have been consumed for actual achievements is a measure of efficiency (*i.e.*, it is an input/output relationship).

6.4 Planned Maintenance Job Costing

6.4.1 Standard Cost Elements of a Planned Maintenance Job

6.4.1.1 Direct Materials (Spare Parts)

Direct materials (spare parts) represent all materials and component parts directly traceable to a planned maintenance job in an economically feasible manner. The direct materials (spare parts) requirements for each planned maintenance job are available on the maintenance schedule for each piece of equipment or on the job specifications document. Such information is initially (or should be) based on a Bill of Materials (BoM). The BoM should allow for normal spoilage of materials if some spoilage is inevitable or related to inherent characteristics of the planned maintenance job.

Hence, the maintenance schedule or the job specification sheet will provide the basis for determining the standard quantities of direct materials that will be reflected in Panel A of the planned maintenance job cost sheet (PMJCS) (see Figure 6.3 and Figure 6.4a: Panel A). These documents also provide the necessary data for the direct materials (spare parts) section of a planned maintenance work order. It is noteworthy to indicate that these documents are prepared at the design stage of the planned maintenance program.

The standard direct materials cost of a planned maintenance job should reflect the quantities allowed as per the bill of materials at prices reflecting normal supply market conditions.

6.4.1.2 Direct Maintenance Labor

Direct maintenance labor represents all labor skills that directly work on a planned maintenance job and their cost is traceable to that job in an economically feasible manner. Planned maintenance direct labor usually comprises a team of several skills needed to ensure the quality and cost effectiveness of the maintenance job. Thus, the mix of the labor skills has to be predetermined and should be reflected in the maintenance schedule or the job specifications document.

The direct maintenance labor information is initially (or should be) based on a job work flow sheet (JWFS). The JWFS is a road map for the maintenance job and provides information about processes to be performed and the labor skill(s) to be applied, the amount of labor time to be utilized under normal conditions. The JWFS should indicate if a certain degree of substitution of labor skills is permissible in order to control the quality and cost of the planned maintenance job. Furthermore, the JWFS should incorporate any inevitable labor downtime due to some inherent characteristics of the maintenance job.

Hence, the maintenance schedule or the job specifications document will provide the basis for determining the standard hours and mix of direct maintenance labor that will be reflected in the Direct Labor section of Panel A of the PMJCS (see Figure 6.4a). The documents also provide the necessary data for the direct labor section of a planned maintenance work order. The standard direct labor cost of a planned maintenance job should reflect the direct labor hours allowed as per the JWFS at wage rates reflecting normal labor market conditions.

6.4.1.3 Support Activities

In addition to direct materials (spare parts) and direct maintenance labor, a planned maintenance job would require the services of support activities in the areas of:

- Design;
- Planning;
- Work order scheduling;
- Dispatching; and
- Follow-up and quality assurance.

Support activities costs represent all planned maintenance costs other than direct materials and direct maintenance labor costs. They can be labeled as planned maintenance overhead costs. These costs are common to all planned maintenance jobs and are not amenable to traceability as direct costs. Hence, the only feasible way to reflect them as part of the costs of a planned maintenance job is through allocation. The question is: on what basis? The following approaches could be followed:

1. Look for a common denominator to serve as a basis for allocating planned maintenance overhead costs such as maintenance job hours or machine hours. However, this approach assumes that all maintenance labor hours or machine hours require the same amount of overhead support. Furthermore, most likely a single basis for overhead allocation may not have any causal relationship with the incurrence of planned maintenance overhead costs. Hence, using a single rate for applying (allocating) these overhead costs to planned maintenance jobs could lead to cost cross-subsidization among maintenance jobs, and eventually would lead to the distortion of planned maintenance costs, making the cost information potentially (if not totally) misleading. In the past, organizations could afford such misleading cost allocations because competitive market forces were not as strong as they are today and because the profitable part(s) of the business outweighed the losing parts. Even not-for-profit organizations did not have an incentive to improve on their costing practices because funding was easier to obtain. Today, the tolerable error margin is narrower and organizations can no longer afford such mistakes and remain competitive or get funded.
2. Since there are different support activities within planned maintenance and since maintenance jobs consume the resources of these activities differently, such differentiation has to be captured in building up a planned maintenance job cost. This issue becomes quite critical when the overhead costs are material (significant) in amount in relation to planned maintenance total costs.

Job No. Scheduled start time:.... Actual start time:	Job Description:	Scheduled finish date: . Actual finish date:
PANEL-A: Standard inputs of direct materials and direct labor (See Figure 6.4a)		PANEL-B: Actual usage of direct materials and direct labor (See Figure 6.4b)
	Variances Summary	
		Amount
	Direct Materials	
	Direct Labor	
	Support Services	

Figure 6.3. Planned maintenance job cost sheet

STANDARD INPUTS													
DIRECT MATERIALS / SPARE PARTS							DIRECT LABOR						
ITEM	ITEM DESCRIPTION	STD	STD	TOTAL	SKILL #	SKILL DESCRIPTION	STD	STD	STD	STD	STD	MIX	AMOUNT
#		PRICE	QTY				RATE	HRS					
TOTAL							TOTAL						
SUPPORT SERVICES				STD	STD	AMOUNT	SUMMARY						
ACTIVITY	ACTIVITY DESCRIPTION	STD	RATE	QTY	COST ELEMENT		AMOUNT						
					DIRECT MATERIALS								
					DIRECT LABOR								
					SUPPORT SERVICES								
					TOTAL								
TOTAL													

Figure 6.4a. Panel-A standard inputs of direct materials and direct labor

ACTUAL USAGE												
DIRECT MATERIALS / SPARE PARTS							DIRECT LABOR					
ITEM #	ITEM DESCRIPTION	STD PRICE	STD QTY	TOTAL	SKILL #	SKILL DESCRIPTION	STD RATE	STD HRS	STD MIX	AMOUNT		
TOTAL							TOTAL					
SUPPORT SERVICES							SUMMARY					
ACTIVITY	ACTIVITY DESCRIPTION	STD RATE		STD QTY		AMOUNT	COST ELEMENT		AMOUNT			
							DIRECT MATERIALS					
							DIRECT LABOR					
							SUPPORT SERVICES					
							TOTAL					
TOTAL												

Figure 6.4b. Panel-B actual usage of direct materials and direct labor

Activity-based costing (ABC) provides the answer, since it does the following:

- Identifies major support activity areas within planned maintenance;
 - For each activity area, identifies cause-and-effect cost driver(s);
 - Develops total budgeted cost (variable and fixed) and total budgeted demand for each activity under normal conditions; and
 - Calculates a predetermined overhead rate per unit of activity for each activity area by dividing total budgeted costs by total budgeted demand.
3. Use the predetermined overhead rates above to apply support overhead costs to planned maintenance jobs on the basis of planned usage of that activity. If a planned maintenance job is not planned to use a given support activity, it receives no allocation of the cost of that activity. Under ABC, allocation of overhead support costs becomes a function of the resources planned to be (or actually) consumed in each activity area. The predetermined overhead rates for the different support activity areas and the planned quantity of support in each activity area are captured to provide the basis for entries in the support services section of Panel A of the PMJCS.

ABC provides appropriate building blocks for reliable maintenance costing as well as a better understanding of the cost structure of a maintenance operation.

6.4.2 Actual Cost Elements of a Planned Maintenance Job

6.4.2.1 Direct Materials (Spare Parts)

The actual consumption of direct materials (spare parts) will be charged in a planned maintenance job cost sheet at standard prices. Why not at actual prices? The reason is that the difference between actual prices and standard prices is a spending variance that is non-controllable by maintenance management. Such a spending variance would be of relevance to procurement management.

A materials (spare parts) requisition form (MRF) will provide documentary evidence about actual consumption of direct materials. The MRF will provide the basis for making direct materials (spare parts) entries in the actual inputs section of the planned maintenance job cost sheet (PMJCS) (see Figure 6.4b, panel B).

6.4.2.2 Direct Maintenance Labor

The actual utilization of direct labor services will be charged in a planned maintenance job cost sheet at standard wage rates. Why not at actual wage rates? The reason is that the difference between the actual labor wage rate and the standard labor wage rate is a spending variance that is non-controllable by maintenance management. Such a spending variance would be of relevance to the human resources department.

Documentary evidence about the actual utilization of direct labor services in terms of hours and mix will be provided by the time ticket for each maintenance worker that provides a record of the elapsed time for each maintenance job in which he worked. These time tickets will provide the input for the direct labor

utilization section of the planned maintenance job cost sheet (PMJCS) (see Figure 6.4b, panel B).

6.4.2.3 Support Services

The actual utilization of planned maintenance activities support will be charged at the predetermined overhead rates. Why at the predetermined overhead rates? The reasons are the following: first, timeliness of the costing of planned maintenance jobs so that their total cost can be determined as soon as they are completed rather than waiting until the end of the fiscal year to determine the “actual” overhead rate of each support activity; and second, avoiding seasonal fluctuations in maintenance costing by basing the overhead rate on estimated costs and volume of support activities under normal conditions.

Documentary evidence about actual utilization of support services by a specific maintenance job will be provided by the actual support activities card. The actual quantity used of a support service will be multiplied by its predetermined overhead rate to arrive at the support services cost to be entered in the support services section of Panel B of the job cost sheet (see Figure 6.4).

6.4.3 Total Cost of a Planned Maintenance Job

After the completion of a planned maintenance job, the totals of the direct materials (spare parts), direct labor, and support services sections will be totaled and the total cost of the job will be summarized in the summary section of Panel B the PMJCS (see Figure 6.4).

6.4.4 Planned Maintenance Job Cost Variances

A cost variance is the difference between an actual cost and a standard cost for an activity or cost object. It could be unfavorable (U) if the actual cost is greater than the standard cost or it could be favorable (F) if actual cost is less than the standard cost. The terms unfavorable or favorable are indicative of the impact of the variance upon the cost of doing business or profitability. It should not be considered as conclusive evidence about the “badness” or “goodness” of managerial performance. For example, the cost variances of a planned maintenance job could be quite favorable because lower quality materials (spare parts) and labor skills were substituted for the quality of materials and labor skills that should have been used.

Conclusive evidence about “goodness” or “badness” of managerial performance or the cost efficiency of a planned maintenance job can be determined only if a significant cost variance is investigated and its causal factors are controllable by management. Possible causal factors could be related to the availability of equipment and facilities for maintenance services, monthly and annual equipment outages or reduction thereof, and quality and safety standards. In short, the financial variances are not an end in themselves but represent a first step toward improving or assuring the cost efficiency of planned maintenance services.

6.4.5 Significant Cost Variances

A significant cost variance is a variance that is worthy of management's attention. The significance level (or control limits) of a variance could be determined quantitatively or judgmentally. Quantitatively, the significance level of a cost variance could be determined by constructing an interval reflecting management's confidence by adding to and subtracting from the mean value of the cost element a multiple (1, 2, or 3) of its standard deviation. The more critical the cost item, the narrower will be the width of the confidence interval.

Judgmentally, the significance level of a cost variance could be determined on the basis of past experience as well as the criticalness of the cost element. For example, if planned maintenance costs are highly sensitive to direct labor cost and management has relevant past experience in controlling that cost, they might set the control limits to $\pm x\%$ of its mean value.

As long as the variance is within its control limits, it does not need to be reported to management. In other words, feedback information is provided to management only if the cost variance falls outside its tolerance limits. This will facilitate management by exception, whereby management's focus is directed toward situations that warrant their attention. In this respect, cost variance reporting falls within the attention-directing role of an accounting information system. However, the behavior of a cost variance over time has to be observed, even though the variance is within its control limits, in order to discern any patterns developing that might result in an out-of-control condition. In such a case management has to be provided with feed forward information to take the appropriate control action(s).

Furthermore, establishing control limits for variances enables the avoidance of information overload since not all variances should be reported to management. The cost variances for direct materials (spare parts), direct labor, and support services could appear individually and in total in the variances section of the planned maintenance job cost sheet (see Figure 6.4).

The variances for planned maintenance cost elements can be computed as follows.

Direct materials (spare parts) efficiency (usage) variance

For each type of material (spare part):

- Standard price per unit $\times$ (actual quantity used *minus* standard quantity allowed for work done).

If the actual usage is greater than the standard allowed usage, the efficiency variance is unfavorable, and it will be favorable if the reverse is true.

Direct maintenance labor efficiency variance

For each labor skill:

- Standard hourly rate $\times$ (actual labor hours used *minus* standard labor hours allowed for work done).

The labor efficiency variance could be decomposed as follows into a labor mix

variance and a labor yield variance to have an idea about how much of the labor efficiency variance is attributable to a change in labor mix and how much is due to a change in labor productivity.

For each labor skill:

- Labor mix variance = standard hourly rate $\times$ (actual labor hours used *minus* actual labor hours used at standard mix)
- Labor productivity variance = standard hourly rate $\times$ (actual labor hours used at standard mix *minus* standard hours allowed for work done at standard mix).

Support services variances for each activity area:

- Standard rate per unit of activity $\times$ (actual units of activity used *minus* standard units of activity allowed for work done).

6.5 Summary and Conclusions

The proposed planned maintenance budgeting and costing framework serves the following purposes:

- Views planned maintenance services as a valuable contributor to the organization's overall cost efficiency and profitability;
- Budgeting for planned maintenance services is driven by the organization's mission statement and business strategy;
- Planned maintenance services will communicate and coordinate its activities with those of all of its internal customers;
- Planned maintenance services will end up with operating and capital budgets adequate for planned service levels;
- Budget performance reports will highlight effectiveness and efficiency of actual maintenance services provided;
- Estimation of standard costs of a planned maintenance job element by element and in total, reflecting an expected level of cost efficiency;
- Accumulation of the actual usage of maintenance resources (inputs) at standard prices, facilitating responsibility accounting for maintenance resources;
- The determination of efficiency variances by cost element and in total;
- This contributes to the efficient utilization of planned maintenance resources;
- Provides timely and reliable cost information to maintenance management;
- Provides timely and reliable information for maintenance modules of ERP systems;
- Facilitates management by exception by directing management's attention to cost variances that are worthy of their attention, providing a sound basis for the appropriate managerial action(s);

- Allows for generation of cost variances under conditions of continuous improvement since the standards for direct materials, direct labor, and support services could be revised to reflect a Kaizen philosophy;
- Provides complete audit trail that facilitates the audit process of planned maintenance costs;
- Provides a framework that could be used in costing unplanned (breakdown) maintenance jobs once the job is defined *ex pos*; and.
- Provides cost information relevant for outsourcing support activities including planned maintenance services so the organization can focus on core activities.

References

- Cokins G (1998) Why is traditional accounting failing managers? *Hosp Mater Manag Q*, November, pp 72–80.
- Duffuaa SO, Raouf, A, Campbell JD (1998) *Planning and Control of Maintenance Systems: Modeling and Analysis*. Wiley, New York.
- Duffuaa SO, Ben Daya M, Al-Sultan K, and Andijani A (2000) A Simulation Model for Maintenance Systems in Saudi Arabia. Final Report, No. AR-16-85.
- Fahy M and O'Brien G (2000) As easy as ABC? It seems not!. *Accountancy Ireland*. 32(1), February.
- Horngren C, Foster G, and Datar S (2007) *Cost Accounting: A Managerial Emphasis*. Prentice-Hall, Englewood Cliffs, NJ.
- Kaplan R. Cooper R (2008) *Activity-Based-Costing: Introduction*. Harvard Business School Press.
- Mirghani MA (1996) Aircraft Maintenance Budgetary and Costing Systems at the Saudi Arabian Airlines: An Integrated Business Approach. *J of Qual in Maint Eng* 2(4): 32–47.
- Mirghani MA (2001) A Framework for Costing Planned Maintenance. *J of Qual in Maint Eng* 7(3): 170–182.
- Mirghani MA (2003) Application and Implementation Issues of a Framework for Costing Planned Maintenance. *J of Qual in Maint Eng* 9(4): 436–449
- Walker M (1998) Attributes or activities? Looking for ABCII. *aust CPA*, 68(9): 26–28.

Simulation Based Approaches for Maintenance Strategies Optimization

Fouad Riane, Olivier Roux, Olivier Basile, and Pierre Dehombreux

7.1 Introduction

Maintenance activities concern the most important assets of firms and can directly impact the competitiveness of companies. Its performance influences the entire production process, from product quality to on-time delivery.

Poor maintenance procedures can cost millions of euros in repairs and can lead to very poor products' quality and substantial production loss while good maintenance practices can cut production costs immensely. Thus maintenance function should no longer be considered as a source of cost but as a critical lever for strategic competitiveness of firms.

Maintenance managers deal with manufacturing systems that are subject to deteriorations and failures. They often have to rethink the way they should deal with maintenance policies and maintenance organization issues. One of their major concerns is the complex decision making problem when they consider the availability aspect as well as the economic issue of their maintenance activities. They are continuously looking for a way to improve the availability of their production machines in order to ensure given production throughputs at the lowest cost.

This decision making problem concerns the allocation of the right budget to the appropriate equipment or component. The objective is to minimize the total expenditure and to maximize the effective availability of production resources.

Different maintenance policies can be applied. Depending on the structure of the production system and its various parameters, managers can define a set of maintenance actions to be executed according to a given schedule. These actions can be derived from different approaches leading to different categories of maintenance strategies: failure based maintenance, use based maintenance, detection based maintenance, condition based maintenance and design-out maintenance (Naert and Van Mol, 2002).

Current maintenance policies are time oriented and are based on reliability models. These models can be classified into two main groups: those developed for non-repairable systems and those considered for repairable systems. Standard models belong to the first family while stochastic processes fit in the second group.

Exponential, Weibull and lognormal distributions are standard reliability models. The stochastic processes can be non-homogenous Poisson processes and generalized renewal processes.

The reliability theory is overviewed by Pham (2003). The theory relative to standard reliability models is largely developed by Ebeling (1997) and Lewis (2004). The mathematical theory relative to stochastic processes including Poisson process, renewal process, markovian process and semi-markovian process is widely discussed by Coccozza-Thivent (1997).

A wide study concerning repairable systems was realized by Ascher and Feingold (1984) who proposed different models. Moreover, literature presents many specific studies relating to reliability models for repairable systems. For example, Calabria and Pulcini (2000) propose two point processes to analyze the failure pattern of a repairable system subject to imperfect maintenance. Coetzee (1997) studies the role of nonhomogeneous Poisson processes (NHPP) models in practical analysis of maintenance failure data. Doyen and Gaudoin (2004) propose a study of age reduction and intensity reduction models. Finally, Yañez *et al.* (2002) present a study of the generalized renewal process.

Reliability models are generally estimated based on small samples. A classical method that can be used to estimate the parameters of reliability models is the maximum likelihood estimation method. This method is largely developed by Meeker and Escobar (1998). Furthermore, on the basis of the likelihood function, one can work out confidence intervals for the estimated parameters. The intervals obtained are called normal-approximation confidence intervals. The confidence intervals are calculated thanks to Monte Carlo simulations and using the variance-covariance matrix.

Once the reliability models are estimated a discrete event simulation model reproducing the dynamic of the system as well as its stochastic behavior can be run in order to validate different maintenance policies and optimize their parameters. The idea is to evaluate the performances of the appropriate strategy before its implementation.

The aim of this chapter is first to provide the reader with the necessary tools allowing reliability estimation and second to calculate the uncertainty affecting reliability parameters estimates with regard to the sample's size. Once the reliability of the system is captured, an integrated framework called OPTIMAIN (2006) will allow maintenance decision makers to design their production system, to model its functioning and to optimize the appropriate maintenance strategies.

7.2 Reliability Models Estimation

7.2.1 Regression and ML Methods

The two methods used traditionally to estimate the parameters of a reliability model from failure times t_i ($i=1 \dots n$) are the regression method and the maximum likelihood method. The parameters estimated by the regression method are determined from the slope and y-intercept axis of the straight line that best fits the data. In practice, we have first to calculate an estimate of the failure function

$F(t_i)=1-R(t_i)$. For that, the estimators generally proposed in literature are the mean rank or median rank estimators.

Next, we have to plot the estimated pairs of points $(t_i, F(t_i))$ on a probability plot which is a graph that corresponds to a linearized model of the function of interest. For example, let us consider the Weibull model characterized by the following reliability function:

$$R(t) = \exp \left[- \left(\frac{t}{\eta} \right)^\beta \right] \quad (7.1)$$

where β is the shape parameter and η is the scale parameter. Applying the logarithm transformation twice to the reliability function, we get the linear expression of this model:

$$\ln \ln \left(\frac{1}{1-F(t)} \right) = \beta \ln(t) - \beta \ln(\eta) \quad (7.2)$$

Therefore, if we plot $\ln(t)$ on the x-axis and $\ln \ln \{1/[1-F(t)]\}$ on the y-axis, then data distributed according to a Weibull model should plot as a straight line. In this case, the parameter β is equal to the slope of the straight line that best fits the data and η is determined from the y-intercept axis of this line such as the one shown in Figure 7.1.

The maximum likelihood method determines the values of the parameters that maximize the probability to observe the data. Therefore, the likelihood function is calculated by the product to observe the failures at each time t_i ($i=1 \dots n$):

$$L(\Theta) = \prod_{i=1}^n f(t_i, \Theta) \quad (7.3)$$

where $\Theta = (\theta_1, \theta_2, \dots, \theta_p)$ is the vector of the model parameters.

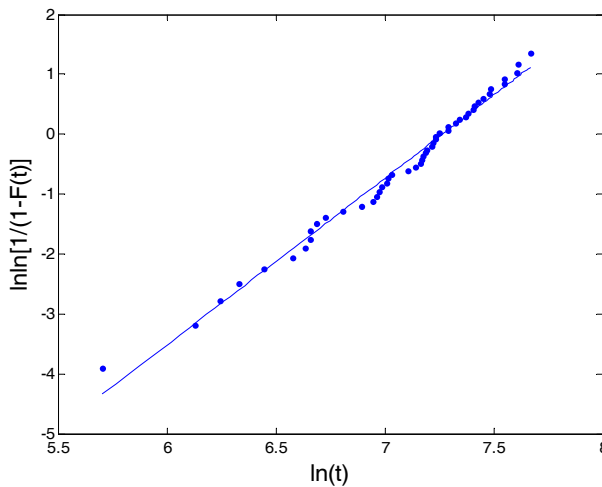


Figure 7.1. Regression method

The expression of the likelihood function for the Weibull model is established as follow:

$$L(\beta, \eta) = \prod_{i=1}^n \frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} \exp \left[- \left(\frac{t_i}{\eta} \right)^{\beta} \right] \quad (7.4)$$

The maximum likelihood estimators maximize the likelihood function and are obtained by equating the first partial derivatives of the function relative to the parameters to zero. Generally we consider the logarithm of the likelihood function. Then, applied to the Weibull distribution we get

$$\begin{cases} \frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n \frac{1}{\beta} - \ln \eta + \ln t_i - \left(\frac{t_i}{\eta} \right)^{\beta} \ln \left(\frac{t_i}{\eta} \right) = 0 \\ \frac{\partial \ln L}{\partial \eta} = \sum_{i=1}^n -\frac{\beta}{\eta} + \frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta} = 0 \end{cases} \quad (7.5)$$

The value of β is estimated by solving the first equation using a numerical method like the Newton-Raphson method. The scale parameter is determined from the second equation that yields

$$\eta = \sqrt[\beta]{\frac{\sum_{i=1}^n t_i}{n}} \quad (7.6)$$

It is obvious that the estimated parameters depend on the data which can be complete or censored. An advantage of the maximum likelihood method is that it accommodates censored data better than regression method. Concerning the estimation accuracy, it depends essentially on the size of datasets: the larger the data size is, the less is the uncertainty. We devote the next section to present some methods for estimating confidence intervals on reliability parameters.

7.2.2 Uncertainty Affecting Reliability Model

To calculate the uncertainty on a reliability parameter θ , one has to compute the limits G_1 and G_2 such that the probability that θ is included in the interval $[G_1, G_2]$ is equal to $1-\alpha$; where α is the confidence level:

$$P(G_1 \leq \theta \leq G_2) = 1 - \alpha \quad (7.7)$$

Therefore, on the basis of θ distribution, we are able to calculate its confidence interval. Literature presents three methods to estimate uncertainty on the basis of:

- The assumption that parameters are normally distributed;
- Likelihood ratio distribution; and
- Simulation (bootstrap methods).

The first method assumes that the parameter θ is distributed according to the normal distribution which mean is equal to the estimated parameter $\hat{\theta}$:

$$\hat{\theta} - z_{1-\alpha/2} se_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{1-\alpha/2} se_{\hat{\theta}} \quad (7.8)$$

where z_p is the p-quantile of the standardized normal distribution and $se_{\hat{\theta}}$ is the standard deviation that can be estimated by the Fisher matrix (Basile *et al.* 2007).

The likelihood ratio is defined as the ratio of the value of likelihood function for a given value of θ relative to the maximum of the likelihood function:

$$r(\theta) = \frac{L(\theta)}{L(\hat{\theta})} \quad (7.9)$$

The determination of a confidence interval on θ is based on the property that the likelihood ratio of the logarithm has asymptotically a chi-square distribution. Then, as represented in Figure 7.2, the uncertainty on θ is given by solving this equation:

$$-2 \ln r(\theta) \geq \chi_{1-\alpha,1}^2 \quad (7.10)$$

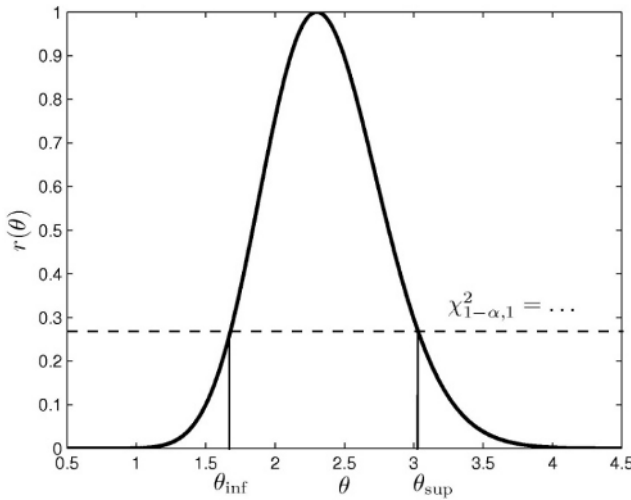


Figure 7.2. Likelihood ratio distribution

Finally, the principle of the simulation based method consists of estimating the θ distribution on the basis of simulated samples of data. For each sample, we estimate the corresponding value of the parameter. Then, we get a sample of different estimated values of θ that allow determining its distribution.

Once the uncertainty on reliability model parameters is fixed, we deduce the uncertainty affecting the reliability law for a given confidence level α as depicted in Figure 7.3 (Basile *et al.* 2007).

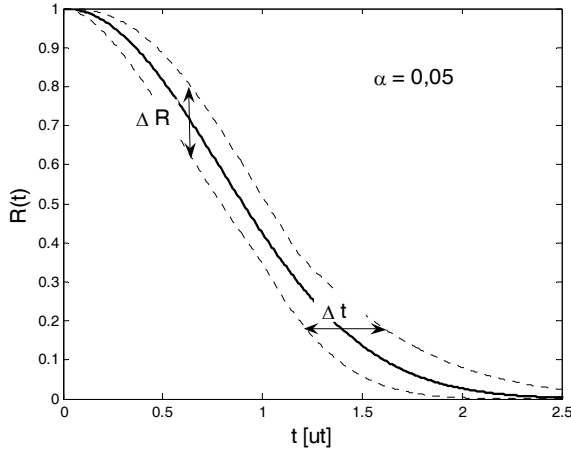


Figure 7.3. Uncertainty affecting the reliability functions

7.3 Maintenance Performance

Besides reliability model estimation, costs and availability are the most important indicators for maintenance performance definition. These indicators can be easily determined. When a system is prone to deterioration, preventive maintenance can reduce maintenance costs and improve the availability of the system. For such systems, the goal of the maintenance manager is to estimate the optimal preventive maintenance schedule.

7.3.1 Availability Model

Availability is defined as the ability of an item (under combined aspects of its reliability, maintainability and maintenance support) to perform its required function at a stated instant of time or over a stated period of time (Rausand and Høyland, 2004). In practice, asymptotic availability is equal to the ratio between the mean time the system operates (Mean Up Time) and the mean time between two failures (MTBF). If we consider the mean down time (MDT) equal to the mean time to repair (MTTR) we get

$$A(t) = \frac{\text{MUT}}{\text{MTBF}} \approx \frac{\text{MUT}}{\text{MUT} + \text{MTTR}} \quad (7.11)$$

Under a preventive maintenance strategy with a periodicity equal to T_p , the mean up time is equal to

$$\text{MUT} | T_p = \int_0^{T_p} R(t) dt \quad (7.12)$$

The plot of the evolution of the function $A(t)$ vs the preventive maintenance periodicity is represented in Figure 7.4. We observe that, for items subject to degradation, there is a maintenance periodicity where availability reaches a maximum.

7.3.2 Costs Model

The determination of the optimal preventive maintenance periodicity considering the costs criteria requires a cost model such as the one presented by Lyonnet (2000). This model distinguishes loss and costs as mentioned in Tables 7.1 and 7.2 where T_a is the production time-stopped, T_i the maintenance time, τ_p is the loss of production per hour, τ_s wages costs per hour, and τ_{am} equipment amortization costs

Maintenance loss and costs are different if the maintenance is corrective or preventive. The average maintenance costs are calculated by the following relationship:

$$\bar{C}_m = \frac{1}{MUT|T_p} \left[\frac{F(T_p)(L_{a,corr} + C_{i,corr})}{(1 - F(T_p))(L_{a,prev} + C_{i,prev})} \right] \quad (7.13)$$

The plot of maintenance costs as a function of the preventive maintenance period is depicted in Figure 7.5. When the hazard function of the system increases with time we can observe a minimum – that obviously depends on maintenance costs and intervention times.

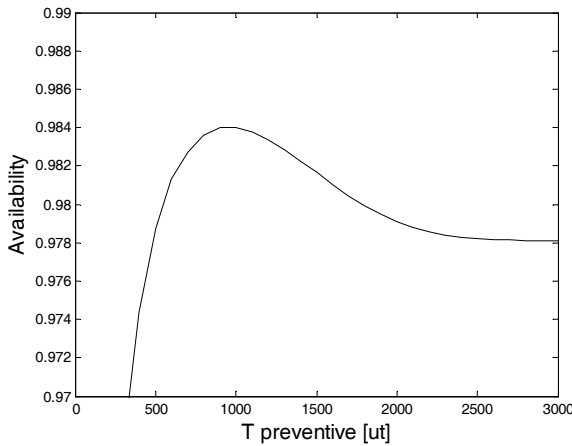


Figure 7.4. Availability in function of maintenance periodicity

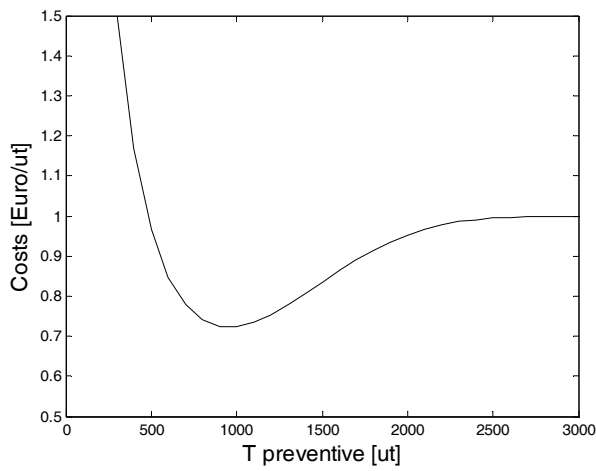


Figure 7.5. Maintenance costs as function of maintenance periodicity

Table 7.1. Loss due to a stop

Production loss	$L_p = \tau_p \times T_a$
Raw material loss	L_{mr}
Failed equipment amortization	L_{ef}
Energy consumed	L_e
Total loss	L_a

Table 7.2. Maintenance operation costs

Wage costs	$C_s = \tau_s \times T_i$
Maintenance equipment amortization	$C_{am} = \tau_{am} \times T_i$
Costs of spare parts stock	C_o
Spare parts costs	C_p
Total of intervention costs	C_i

The determination of ‘optimal’ maintenance periodicities (optimal periodicity is different considering availability or costs criteria) can be estimated analytically for a simple system. For a complex system, only simulation allows the determination of these periodicities. This issue is addressed in the remainder of this chapter.

7.4 Simulation Based Maintenance Framework

7.4.1 Toward a Unified Framework

In the manufacturing context, many problems are of such complexity that managers have to build their decisions based on aid support systems. These systems should be adequate to provide satisfactory answers to their strategic, tactical or operational questions. Most of these problems can be modeled as complex discrete systems, whose related decision making requirements are approached using simulation techniques.

We are concerned with the development of a simulation based decision support approach that can operate for designing maintenance strategies for complex production systems. In particular, we address the development of a unified, graphical framework that makes it possible for the decision maker to, first, understand and to model the dynamics of the considered system and, second, to design and to optimize the appropriate maintenance policy. A key element of such a framework is the development of a graphical language that enables the automatic code generation for simulation purposes and optimization analysis.

To handle the elaboration of maintenance strategies for complex system one has to specify the structure of the system, its logical organization, its maintenance strategies and its decision parameters (preventive and corrective costs, preventive periodicity, *etc.*). The models of the system as well as the data will be developed following a unified methodology that leads to the specifications of the real system. Then, detailed level analysis that integrates different scenarios’ comparison is derived. The methodology uses a systemic and hierarchical approach.

The architecture of the open environment that includes concepts, methodologies, languages, and solving engines and is used to support the maintenance framework is depicted in Figure 7.6. The overall objective of such architecture is the integration of several tools in a unified manner. The unification aspect of the framework relies on the application of the same set of concepts at different levels from the modeling methodology to the solving engines. This framework is actually supported by a tool, named OPTIMAIN, developed in the context of a research program funded thanks to the support of the Walloon Region of Belgium.

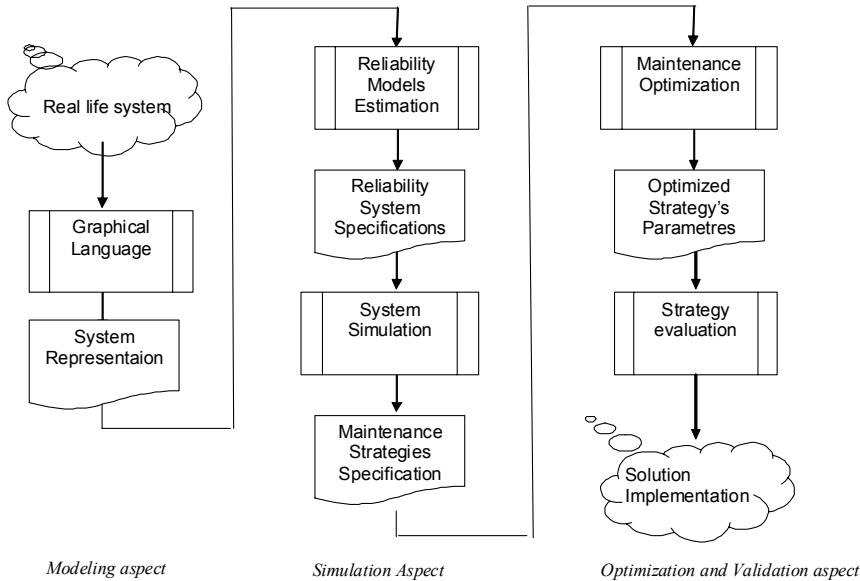


Figure 7.6. A unified approach for solving decision making problems

7.4.2 Maintenance Strategies

A lot of work has been done in the field of maintenance policies of deteriorating equipment. A maintenance strategy may be defined as a decision rule which establishes the sequence of maintenance actions to be undertaken according to the degradation level of the system and with regard to the acceptable exploitation thresholds. Each maintenance action consists of maintaining or restoring the system in a specified state using the appropriate resources. A cost and duration are incurred to execute each maintenance action.

Different maintenance strategies can be encountered in the literature (Lyonnet, 2000; Cho and Parlar, 1991; Nakagawa, 1979; Pierskalla and Voelker, 1976; Sherif and Smith, 1981). They concern the replacement of systems subject to random failures and whose states are known at all times. All the studied policies are governed by analytical models that make it possible to evaluate over an infinite horizon the associated performances under a series of hypothesis (Ait-Kadi *et al.* 2002). These strategies differ from each other by the nature and the action sequel that they suggest, by the selected performance criteria, by the deterministic or stochastic character of the parameters that they take into account, by the fact that the system is considered as a sole entity or as a system constituted of many components which state may be known at all time or after inspection, *etc.*

Using an analytical formulation, one can model the considered maintenance strategy using its characteristic parameters and decision variables to describe the technical as well as the economical objectives to optimize. If one succeeds to solve to optimality such analytical models, he can establish the existence and the

uniqueness conditions of an optimal strategy. He can also derive sensibility analysis.

The major inconvenience of such approaches is that one ends up with difficult models that are complex to solve, especially if one wants to consider other factors that have significant impact on the system's behavior. This fact has brought us to explore simulation's possibilities in order to handle the situation and efficiently evaluate maintenance strategies. For illustrative purposes, we develop a simulation model for a different version of the Modified Bloc Replacement Policy (MPRP) using RAO simulation language (Artiba *et al.* 1998).

Let us consider a single component system that is subject to random failures. Each time the system breaks down we replace it with a new one. On the other hand, when performing preventive maintenance, we replace the system only if its age is greater than a given threshold b . These preventive actions are scheduled at given dates kT ($k=1,2,3, \dots$). The scheme described above is called the Modified Bloc Replacement Policy. It was suggested in order to improve the Bloc Replacement Strategy. The Achilles heel of this latter strategy lies in the fact that components are changed preventively even if they are almost new.

A maintenance manager who chooses to implement this policy looks for the replacement period T and the age threshold b that maximize the steady-state availability of the system or that minimize the expected cost per unit time.

We can improve this strategy in order to integrate risk analysis as a condition in order to activate maintenance actions. This strategy follows the same scheme as MBRP. The difference lies in the preconditions to be satisfied in order to fulfill a preventive action. The component life cycle is punctuated by a sequence of failure events and related preventive actions. The single component system swings between a broken down state and a ready state. The state transition is dictated by the value of the generated moment of breakdown (Riane *et al.* 2004). Failure occurs if this value is less or equal to the remaining time-to-preventive action. Otherwise, the system is still working but can be maintained preventively.

The preventive action is realized only if the value of risk, denoted by PD, is lower than the level accepted by the decision maker (captured in the model by accepted risk level). A mathematical analysis is then necessary to evaluate the value risk based on the computing of the failure's probability between kT and $(k+1)T$, knowing that the component has survived until kT . This involves evaluating integrals of the density function f for kT and $(k+1)T$. We use a numerical approach when density functions are not suitable for integral computation which is the case of the normal distribution. The model's diagram is depicted in Figure 7.7 thanks to ALIX modeling formalism that is suitable for use with RAO simulator (Pichel *et al.* 2003).

The simulation language RAO¹ is based on the RAO (resources - actions - operations) method, which uses modified production rules for describing complex discrete systems (CDS) and processes. Production equipment can be modeled as a complex discrete system using resources and performing operations. The resources are depicted in a database. The set of operations (actions) fulfilled by the resources

¹ Developed within the "Centre de Recherches et d'Etudes en Gestion Industrielle" (CREGI)

are defined by the modified production rules described in a knowledge base (Figure 7.8). Unlike traditional production rules, the modified ones make it possible to describe the dynamics of the system thanks to the temporal specifications and dependence relations between activities.

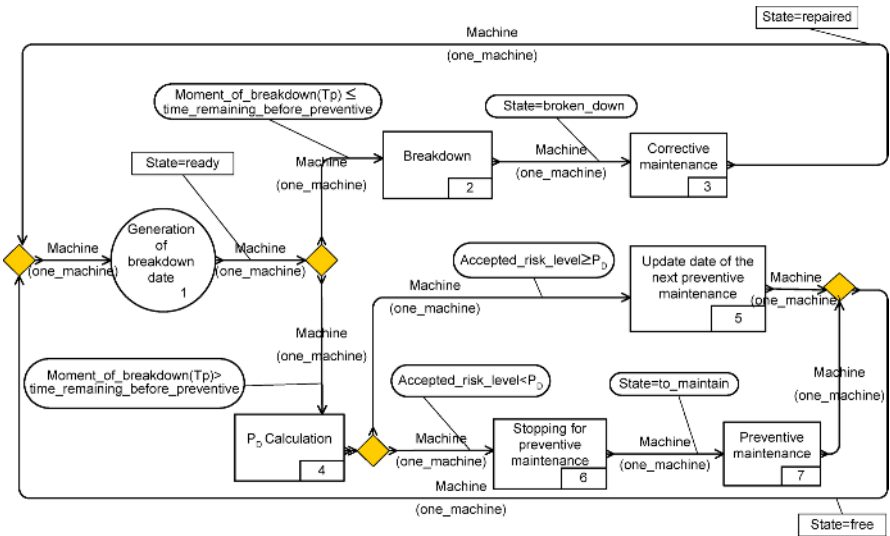


Figure 7.7. Simulation model's flow chart of the block type strategy

To reproduce the functioning process of a complex discrete system in the RAO simulator, the modeler has to describe the concurrency of the irregular events and the way they influence the realization of the different actions.

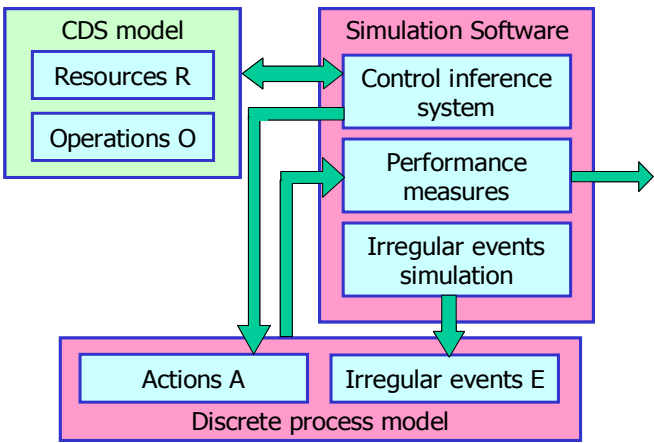


Figure 7.8. Principles of RAO running

A user-friendly interface was developed to support the framework (Figure 7.9). It allows the user to model its system easily using block diagram formalism. For each system's component, the user specifies the appropriate data and the system characteristics are saved in an XML format.

The results of simulation can be presented in different ways using Gantt diagrams, or evolution curves and trace files (Figure 7.10).

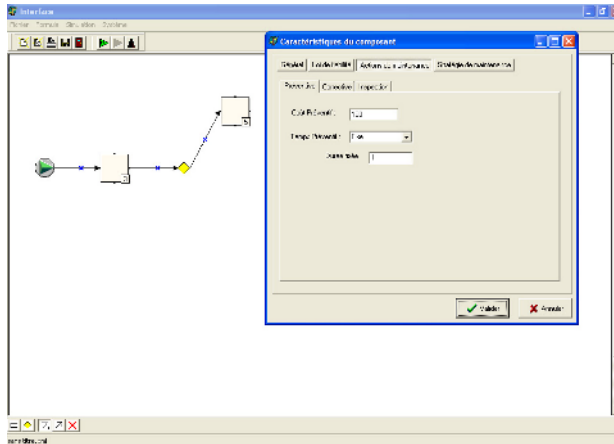


Figure 7.9. Simulation maintenance interface

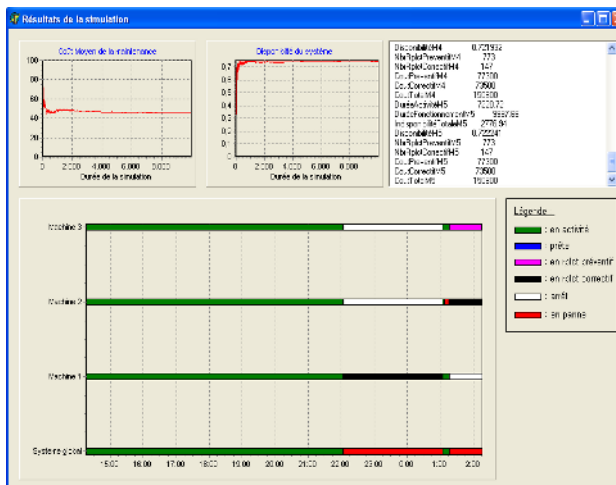


Figure 7.10. Simulation results presentation

7.4.3 Uncertainty Affecting Maintenance Performances

The simulation of the dynamics of the system is possible thanks to the use of RAO simulator. All the necessary models are generated automatically by the framework.

Unfortunately, simulation tells us that identifying the optimal maintenance periodicity is not obvious. The reason is that events occur randomly and then performance indicators are affected by uncertainty as represented in Figures 7.11 and 7.12.

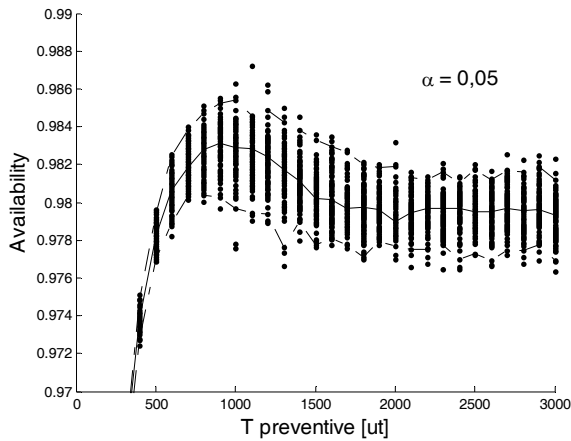


Figure 7.11. Uncertainty on availability

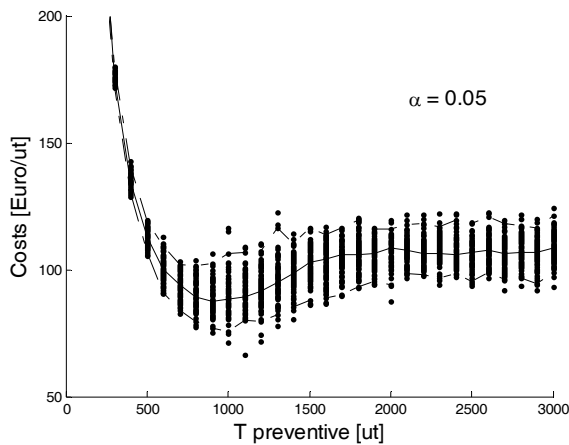


Figure 7.12. Uncertainty on maintenance costs

As a consequence, the maintenance manager is not able to assert precise maintenance performance variables; but he can specify a confidence level from the

study of indicators spread as represented in Figures 7.13 and 7.14. In practice, it is possible to announce that availability and costs will not be exceeding given levels in x % of the cases.

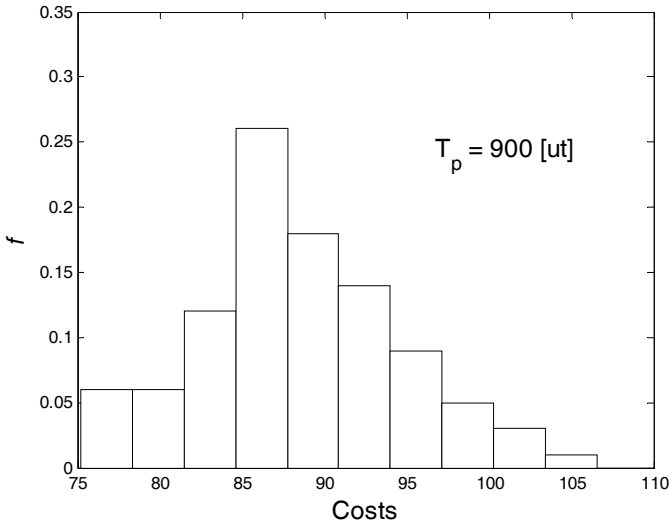


Figure 7.13. Costs spread for $T_p=900$ ut

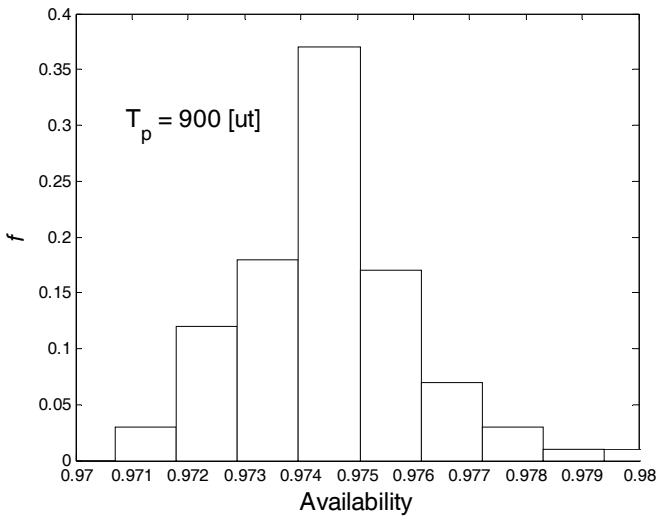


Figure 7.14. Availability spread for $T_p=900$ ut

7.5 A Case Study

In the following we shall use a real life based example to illustrate the different steps needed to deploy the framework discussed above.

The example we are covering is a series-parallel system composed of four main machines named Poste 1, Poste 2, Poste 3a and Poste 4a. The system has four failures modes, as depicted in Figure 7.15. Machines Poste 3b and Poste 4b are redundant to increase the reliability of the process.

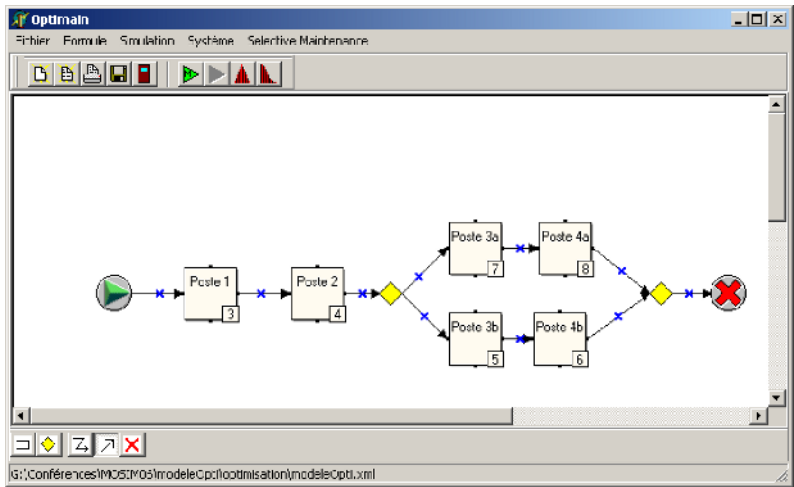


Figure 7.15. The multi-component hybrid system studied

The system needs to be characterized by its economic data and its reliability parameters. The economic data are summarized in Table 7.3 and refer to the replacement fixed costs and the associated durations of corrective and preventive maintenance actions.

Every process stop triggers an incremental loss of productivity rate of 4200 €/h and a manpower cost rate of 230 €/h.

Table 7.3. The economic data

Machine	C_p [€]	C_c [€]	Maintenance actions' duration	
			Preventive [h]	Corrective [h]
Poste 1	5233	21455	6	12
Poste 2	1248	6241	30	48
Poste 3	9358	11697	24	24
Poste 4	3027	9231	36	36

Machines can break down. We assume that the density functions of the machines' failures follow a Weibull distribution with two parameter η and β . Their estimated values were calculated based on historical failure data and are summarized in Table 7.4. We used the methods described before to obtain the most accurate estimation.

Since we consider a multi-component system, we assume that a failed component is detected only when the system stops. The replaced components are considered to be as good as new.

Once the reliability of the system is modeled, we can run simulations to compute the mean up time (MUT) for each machine and for the whole system. Simulation also allows the derivation of the reliability function of the system (Figure 7.16).

Table 7.4. The Weibull parameters and MUT estimates

Machine	MUT [h]	η	β
Poste 1	9216	10364	2.3
Poste 2	8856	9883	3.1
Poste 3	2815	3102	3.8
Poste 4	5572	6043	1.3
System	2837		

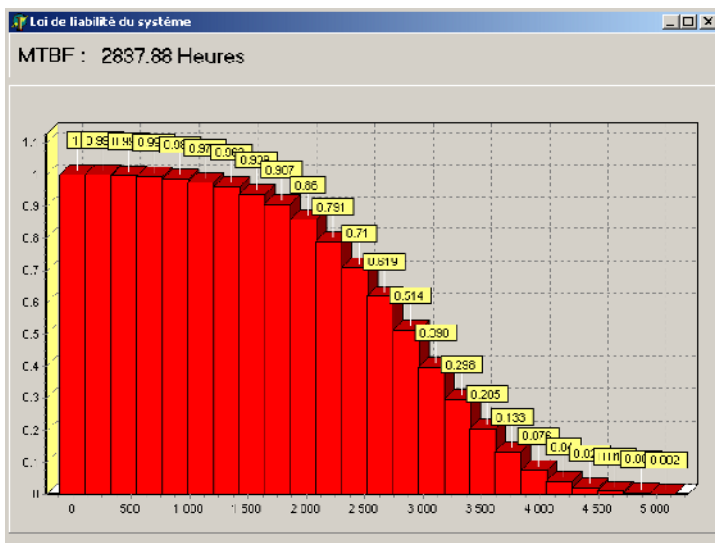


Figure 7.16. The computed system reliability function

We have succeeded in modeling the stochastic behavior of the systems. We now need to establish the maintenance strategy that either minimizes the total cost

or maximizes the system's availability. This optimization phase consists of specifying maintenance policies for each machine and then computing the optimized values for these policies parameters.

Besides the Modified Bloc Replacement Policy described before, two other basic policies are tested in this study. The Age Replacement Policy (ARP) introduced by Barlow and Proschan (1965), which suggests replacing the item at failure or at given age T , whichever occurs first. Only new items are used to perform replacement.

Barlow and Proschan also consider the Block Replacement Policy (BRP) where the replacements are undertaken at KT periods ($K=1, 2, 3 \dots$) or at failure. Only new items are used to perform replacement.

Simulation models of these policies, shown in Figures 7.17 and 7.18, are implemented in the OPTIMAIN framework.

OPTIMAIN allows easy comparison of the performances of different policies when the maintenance periodicity of each component is fixed. If we consider that components are stopped for preventive maintenance each MUP period, we obtain the performances depicted in Table 7.5. We can also compare the results to those obtained where no preventive maintenance is considered.

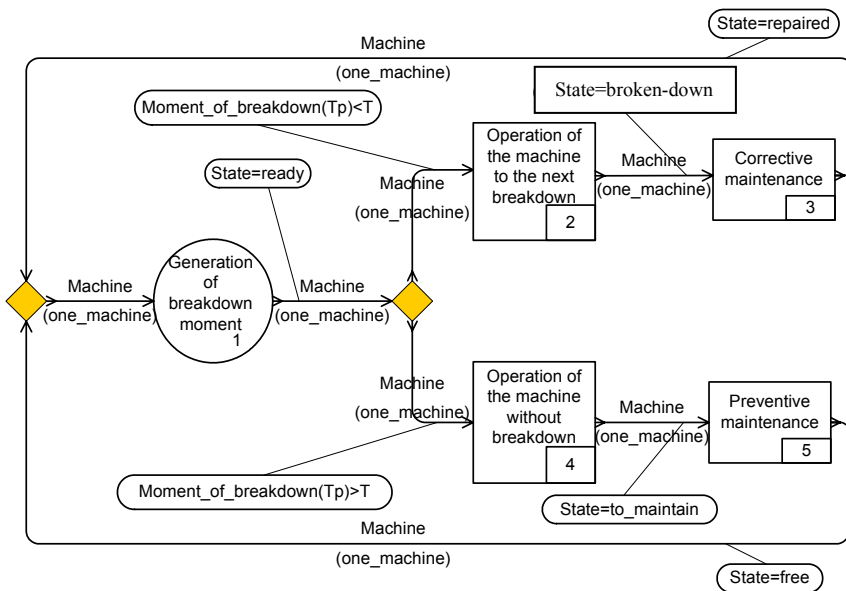


Figure 7.17. Simulation model of the age type strategy flow chart

The results obtained are not convincing and push us to conduct a real optimization. The optimization phase is intended to determine the optimal periodicity T^* which minimizes the total cost or maximizes the stationary availability of the system subjected, respectively, to the age replacement policy (ARP), the block replacement policy (BRP), or the modified bloc replacement policy (MBRP).

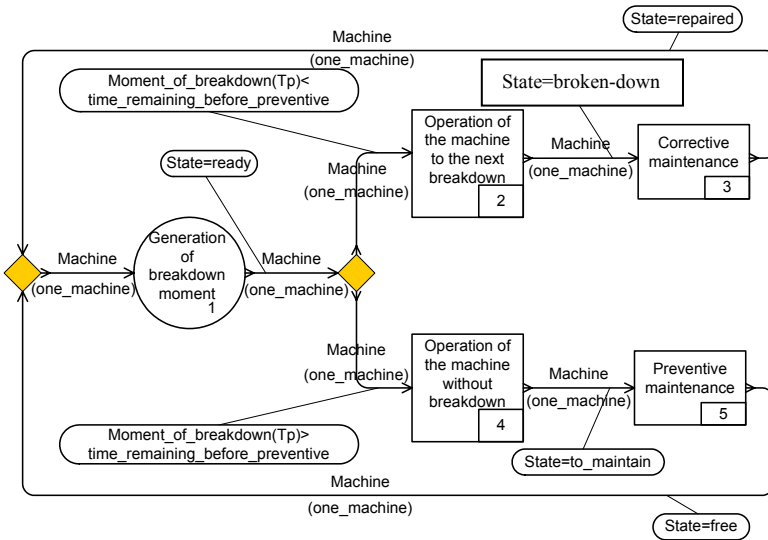


Figure 7.18. Simulation model of the block type strategy flow chart

Table 7.5. Performances of maintenance strategies with $T=MUT$

Policy	Total cost	Standard deviation	Availability	Standard deviation
MUP	[€/h]	[€/h]	[%]	[%]
Corrective	94.17	2.99	98.19	0.06
BRP	158.93	3.21	97.24	0.05
ARP	94.24	2.09	98.24	0.05
MBRP	117.36	2.94	98.05	0.06

We have implemented an optimization algorithm based on the Nelder-Mead method (Nelder and Mead, 1965). It is a local search optimization algorithm commonly used for its simplicity of programming, its low use of memory (few variables), and its reasonable computing time.

The optimized results are presented in Table 7.6. We notice that the BRP and MBRP policies take benefit from optimization and gain in performance accuracy even if the ARP policy states the best strategy for this example.

Table 7.6. Performances of the optimized strategies.

Policy	Total cost	Standard deviation	Availability	Standard deviation
	[€/h]	[€/h]	[%]	[%]
BRP	103.9	2.49	98.44	0.06
ARP	93.62	1.62	98.31	0.04
MBRP	94.68	2.51	98.51	0.06

7.6 Conclusion

Maintenance managers are concerned with the optimization of the availability of their production machines to ensure production throughputs at lowest expenditure. They need a clear process to choose an optimal maintenance strategy for their complex systems whose operating characteristics deteriorate with use and whose lifetime and repair time are random.

Maintenance strategies evaluation and optimization taking into account all considerations and factors that have a significant impact on system's control and on its performance can lead to complex analytical models or even sometimes models are difficult to develop. This observation has led us to explore numerical simulations potential combined with optimization algorithms to evaluate and optimize the performance of maintenance strategies.

We have developed a modeling approach that is supported by a framework called OPTIMAIN. It combines optimization and simulation and makes it possible to capture the dynamic behavior as well as the reliability of multi-component systems. The use of simulation provides an easy evaluation for a maintenance system's performance in terms of availability and average cost per unit time.

OPTIMAIN framework addresses all the aspects of reliability estimation, stochastic and dynamic modeling, maintenance policies evaluation and optimization. It allows managers to accurately build sound decisions.

The integration of optimization techniques helps to find the optimal values of various parameters for appropriate implementation of these strategies. Nevertheless, the use of such techniques is based on reliability models that need parameter estimation. They are also subject to uncertainties that could affect the results and induce errors in maintenance parameter optimization.

References

- Ait-Kadi D, Jamali MA, Artiba A (2002) Optimal periodic replacement strategies using new and used items. *J of Decis Syst* 12: 67–77.

- Artiba A, Emelyanov VV and Iassinovski SI (1998) Introduction to intelligent simulation: The RAO language. Kluwer Academic Publishers.
- Ascher H and Feingold H (1984) Repairable Systems Reliability – modeling, inference, misconceptions, and their causes. New York: Marcel Dekker.
- Barlow R and Proschan F (1965) Mathematical Theory of Reliability, John Wiley & Sons, New York.
- Basile O, Dehombreux P, Riane F (2007) Precision of Reliability Models for Repairable Systems. *J of Qual in Maint Eng* 13: 137–151.
- Calabria R and Pulcini G (2000) Inference and test in modeling the failure/repair process of repairable mechanical equipments. *Reliab Eng and Syst Saf* 67: 41–53.
- Cho DI and Parlar M (1991) A survey of maintenance models for multiunit systems. *Eur Jof Oper Res* 51: 1–23.
- Cocozza-Thivent C (1997) *Processus stochastiques et fiabilité des systèmes*. Springer.
- Coetzee JL (1997) The role of NHPP models in the practical analysis of maintenance failure data. *Reliab Eng and Syst Saf* 56: 161–168.
- Doyen L and Gaudoin O (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliab Eng and Syst Saf* 84: 45–56.
- Ebeling CE (1997) *An Introduction to Reliability and Maintainability Engineering*. McGraw-Hill.
- Fleurquin G, Riane F, Roland B, Dehombreux P, Hou G (2006) OPTIMAIN : Plateforme d'élaboration de stratégies de maintenance de systèmes complexes. Actes de la 6e conférence francophone de modélisation et simulation, MOSIM'06, Rabat, Maroc.
- Lewis EE (2004) *Introduction to Reliability Engineering*. McGraw-Hill, 2nd edition.
- Lyonnet P (2000) *La Maintenance, mathématiques et méthodes*. Tec & Doc.
- Meeker WQ and Escobar LA (1998) *Statistical Methods for Reliability Data*. Wiley series in probability and statistics.
- Naert A and Van Mol T (2002) Development and Implementation of a Total Productivity Improvement Concept, Centrum voor Industrieel Beleid, KULeuven.
- Nakagawa T (1979) Optimum policies when preventive maintenance is imperfect. *IEEE Trans on Reliab R-28*: 331–332.
- Nelder JA and Mead R (1965), A simplex method for function minimization. *The Comput J* 7: 308–313.
- Pham H (2003) *Handbook of Reliability Engineering*. Springer.
- Pichel D, Dontaine A, Iassinovski S, Artiba A and Fagnart C (2003) ALIX: une méthode de modélisation des flux d'un système de production, Conf. CPI 2003, Maroc.
- Pierskalla WP and Voelker JA (1976) A survey of maintenance models: The control and surveillance of deteriorating systems. *Nav Res Logist Q* 23: 353–388.
- Rausand M and Høyland A (2004) *System Reliability Theory, Models, Statistical Methods, and Applications*. Wiley Series in Probability and Statistics.
- Riane F, Roux O, Dehombreux P, and Bardey D (2004) Maintenance Strategies' development from a risk aversion point of view, Intelligent Maintenance Systems International Conference (IMS'04), Arles (France), 14–17 July.
- Sherif YS and Smith ML (1981) Optimal maintenance models for systems subject to failure – a review. *Nav Res Logist Quarterl* 28: 47–74.
- Yañez M, Joglar M and Modarres M (2002) Generalized renewal process for analysis of repairable systems with limited failure experience. *Reliab Eng and Syst Saf* 77: 167–180.

Maintenance Planning and Scheduling

Maintenance Forecasting and Capacity Planning

Hesham K. Al-Fares and Salih O. Duffuaa

8.1 Introduction

Carrying out an effective maintenance operation requires efficient planning of maintenance activities and resources. Since planning is performed in order to prepare for future maintenance tasks, it must be based on good estimates of the future maintenance workload. The maintenance workload consists of two major components: (1) scheduled and planned preventive maintenance, including planned overhauls and shutdowns, and (2) emergency or breakdown failure maintenance. The first component is the deterministic part of the maintenance workload. The second component is the stochastic part that depends on the probabilistic failure pattern, and it is the main cause of uncertainty in maintenance forecasting and capacity planning.

Estimates of the future maintenance workload are obtained by forecasting, which can be simply defined as predicting the future. Clearly, good forecasts of the maintenance workload are needed in order to plan well for maintenance resources. In terms of the time horizon, forecasts are typically classified into three main types: (1) short-term ranging from days to weeks, (2) intermediate-term ranging from weeks to months, and (3) long-term ranging from months to years. Long-term forecasts are usually associated with long-range maintenance capacity planning.

The main objective in capacity planning is to assign fixed maintenance capacity (resources) to meet fluctuating maintenance workload in order to achieve the best utilization of limited resources. Maintenance capacity planning determines the appropriate level and workload assignment of different maintenance resources in each planning period. Examples of maintenance resources include spare parts, manpower of different skills (craftsmen), tools, instruments, time, and money. For each planning period, capacity planning decisions include the number of employees, the backlog level, overtime workload, and subcontract workload. Proper allocation of the various maintenance resources to meet a probabilistic fluctuating workload is a complex and important practical problem. In order to

solve this problem optimally, we have to balance simultaneously the cost and availability of all applicable maintenance resources. A variety of capacity-planning techniques are used for handling this complex problem.

This chapter presents the main concepts and tools of maintenance workload forecasting and capacity planning. Section 8.2 provides a brief introduction to forecasting. Section 8.3 describes qualitative or subjective forecasting techniques. Section 8.4 presents quantitative or objective forecasting models. Section 8.5 covers model evaluation and error analysis. Section 8.6 presents different approaches to maintenance workload forecasting. Section 8.7 outlines the problem of capacity planning in maintenance. Sections 8.8 and 8.9 respectively describe deterministic and stochastic techniques for capacity planning. Finally, Section 8.10 gives a brief summary of this chapter.

8.2 Forecasting Basics

Forecasting techniques are generally classified into two main types: qualitative and quantitative. Qualitative (subjective) techniques are naturally used in the absence of historical data (*e.g.*, for new machines or products), and they are based on personal or expert judgment. On the other hand, quantitative (objective) techniques are used with existing numerical data (*e.g.*, for old machines and products), and they are based on mathematical and statistical methods.

Qualitative forecasting techniques include historical analogy, sales force composites, customer surveys, executive opinions, and the Delphi method. Quantitative techniques are classified into two types: (1) growth or time-series models that use only past values of the variable being predicted, and (2) causal or predictor-variable models that use data of other (predictor) variables.

Nahmias (2005) makes the following observations about forecasts: (1) forecasts are usually not exact, (2) a forecast range is better than a single number, (3) aggregate forecasts are more accurate than single-item forecasts, (4) accuracy of forecasts is higher with shorter time horizons, and (5) forecasts should not ignore known and relevant information. To choose a forecasting technique, the main criteria include: (1) objective of the forecast, (2) time horizon for the forecast, and (3) data availability for the given technique. In order to develop a quantitative forecasting model, the steps below should be followed:

1. Define the variable to be predicted, and identify possible cause-effect relationships and associated predictor variables;
2. Collect and validate available data for errors and outliers;
3. Plot the data over time, and look for major patterns including stationarity, trends, and seasonality;
4. Propose several forecasting models, and determine the parameters and forecasts of each model;
5. Use error analysis to test and validate the models and select the best one; and
6. Refine the selected model and try to improve its performance.

Quantitative forecasting techniques are classified into time-series and causal models. They aim to identify, from past values, the main patterns that will continue in the future. The most frequent patterns, illustrated in Figure 8.1, include the following:

1. Stationary: level or constant demand;
2. Growth or trend: long-term pattern of growth or decline;
3. Seasonality: cyclic pattern repeating itself at fixed intervals; and
4. Economic cycles: similar to seasonality, but length and magnitude of cycle may vary.

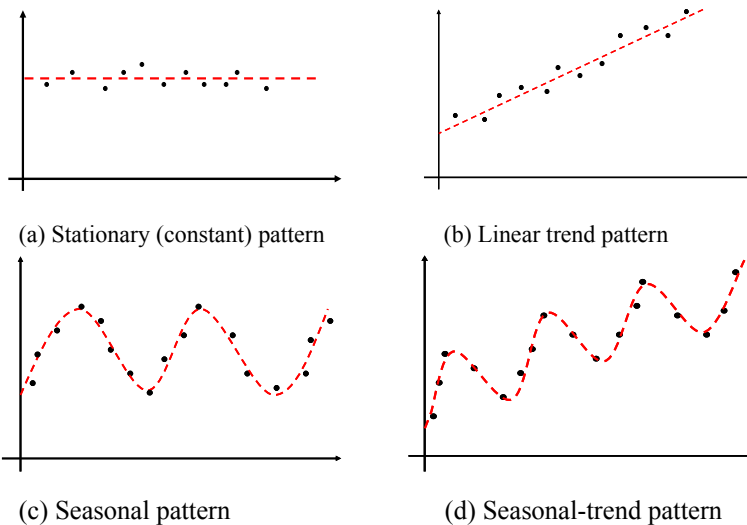


Figure 8.1. Major patterns identified in quantitative forecasting techniques

8.3 Qualitative Forecasting Techniques

Qualitative or subjective forecasting is used in any case where quantitative forecasting techniques are not applicable. Such cases include non-existence, non-availability, non-reliability, and confidentiality of data. Qualitative forecasting is also used when the forecasting horizon is very long, *e.g.*, 20 years or more, such that quantitative forecasting techniques become unreliable. In the absence of numerical data, good qualitative forecasts can still be obtained by systematically soliciting the best subjective estimates of the experts in the given field. For maintenance requirements of new plants or equipment, qualitative forecasting techniques include benchmarking with similar plants and referring to the maintenance instructions provided by the equipment manufacturers. Nahmias (2005) identifies four types of subjective forecasting techniques:

1. Sales force composite: each member of the sales force submits a forecast for items he or she sells, and then the management consolidates.
2. Customer surveys: collect direct customer input; must be carefully designed to find future trends and shifting preferences.
3. Executive opinions: forecasts are provided by management team members from marketing, finance, and production.
4. The Delphi method: a group of experts respond individually to a questionnaire, providing forecasts and justifications. Results are combined, summarized, and returned to experts to revise. The process is repeated until consensus is reached

The most sophisticated technique for qualitative forecasting is the Delphi method, which will be presented in the next section.

8.3.1 The Delphi Method

The Delphi method is a systematic interactive qualitative forecasting technique for obtaining forecasts from a panel of independent experts. The experts are carefully selected and usually consulted using structured questionnaires that are conducted in two or more rounds. At the end of each round, an anonymous summary of the experts' latest forecasts as well as the reasons they provided for their judgments is provided to the experts by a facilitator. The participants are encouraged to revise their earlier answers in light of the replies of other members of the group.

It is believed that during this process the variations in the answers will gradually diminish and that the group will converge towards a consensus. The process is terminated after a pre-defined stopping criterion (*e.g.*, number of rounds, achievement of consensus, and stability of results). According to Rowe and Wright (2001), the mean or median scores of the last round determine the final estimates. The Delphi method was developed in the 1950s by the RAND Corporation in Santa Monica, California. The following steps may be used to implement a Delphi forecasting process:

1. Form the Delphi team to conduct the project;
2. Select the panel of experts;
3. Develop the Delphi questionnaire for the first round;
4. Test and validate the questionnaire for proper design and wording;
5. Send the first survey to the panel;
6. Analyze the first round responses;
7. Prepare the next round questionnaire and possible consensus tests;
8. Send the next round questionnaire to the experts;
9. Analyze responses to the questionnaire (steps 7 through 9 are repeated until the stopping criterion is satisfied); and
10. Prepare the report with results, analysis, and recommendations.

The Delphi method is based on the following assumptions: (1) well-informed individuals using their insight and experience can predict the future better than theoretical models, (2) the problem under consideration is very complex, (3) there

is no history of sustained communication among participating experts, and (4) exchange of ideas is impossible or impractical.

The strengths of the Delphi method include: (1) it achieves rapid consensus, (2) participants can be anywhere in the world, (3) it can cover a wide range of expertise, and (4) it avoids groupthink. The limitations of the method include: (1) it neglects cross impact, (2) it does not cope well with paradigm shifts, and (3) its success depends on the quality of the experts.

The Delphi method can be applied in maintenance in several areas, including determining time standards and preventive maintenance time intervals, as well as estimating the remaining useful life of equipment.

8.4 Quantitative Forecasting Techniques

Quantitative or objective forecasting techniques are presented in this section. These models are based on the availability of historical data, and are usually classified into time-series and causal models. A time series is a set of values of the variable being predicted at discrete points in time. Time-series models are considered naïve because they require only past values of the variable being predicted. Causal models assume that other predictor variables exist that can provide a functional relationship to predict the variable being forecasted. For example, the age of given machine equipment may help in predicting the frequency of failures. The models presented here include methods for stationary, linear, and seasonal data.

8.4.1 Simple Moving Averages

This type of forecast is used for stationary time series, which is composed of a constant term plus random fluctuation. An example of this could be the load exerted on an electronic component. Mathematically, this can be represented as

$$D_t = \mu + \varepsilon_t \quad (8.1)$$

where

$$\begin{aligned} D_t &= \text{demand at time period } t, \\ \mu &= \text{a constant mean of the series,} \\ \varepsilon_t &= \text{error at time } t; \text{ a random variable with mean 0 and variance } \sigma^2. \end{aligned}$$

Obviously, our forecast of future demand should be our best estimate of the parameter μ . Let us assume that all N previous observations are assumed to be equally important, *i.e.*, equally weighted. If we use the least-squares method, then we look for the value of μ that minimizes the sum of squared errors (SSE):

$$SSE = \sum_{t=1}^N (D_t - \mu)^2 \quad (8.2)$$

When we differentiate Equation 8.2 with respect to μ and equate the result to zero, we obtain the optimum value of μ as our forecast given by

$$F_t = \frac{\sum_{i=1}^N D_{t-i}}{N} \quad (8.3)$$

where

$$F_t = \text{forecast for time periods } t, \dots, \infty$$

Since F_t is the average of the last N actual observations (periods $t-1, \dots, t-N$), it is called a simple N -period moving average, or a moving average of order N .

If simple moving average Equation 8.3 is used with a perfectly linear data of the form $D_t = a + bt$, then there will be an error that depends on the slope b and the number of points included in the moving average N . Specifically, the forecast will underestimate or lag behind the actual demand by

$$\varepsilon_t = D_t - F_t = \frac{(N+1)b}{2} \quad (8.4)$$

Example 8.1: The breakdown maintenance load in man-hours for the last 5 months is given as

t	1	2	3	4	5
D_t	800	600	900	700	600

Forecast the maintenance load for period 6 using a 3-month moving average.

The forecasted load for month 6 and all future months is

$$F_6 = \frac{900 + 700 + 600}{3} = 733.33$$

8.4.2 Weighted Moving Average

In simple moving average, an equal weight is given to all n data points. Since individual weight is equal to $1/n$, then sum of the weights is $n(1/n) = 1$. Naturally, one would expect that the more recent data points have more forecasting value than older data points. Therefore, the simple moving average method is sometimes modified by including weights that decrease with the age of the data. The forecasting model becomes

$$F_t = \sum_{i=1}^N w_i D_{t-i} \quad (8.5)$$

where

w_i = weight of the i th observation in the N -period moving average

$$\sum_{i=1}^N w_i = 1 \quad (8.6)$$

The values of w_t must be non-decreasing with respect to t . These values can be empirically determined based on error analysis, or subjectively estimated based on experience, hence combining qualitative and quantitative forecasting approaches.

Example 8.2: Using the maintenance load values of Example 8.1, assume that each observation should weigh twice as much as the previous observation. Forecast the load for month 6 using a 3-period weighted moving average:

$$w_2 = 2w_1$$

$$w_3 = 2w_2$$

$$w_1 + w_2 + w_3 = 1$$

Solving the 3×3 system gives

$$w_1 = 1/7$$

$$w_2 = 2/7$$

$$w_3 = 4/7$$

The forecasted load for month 6 and all future months is

$$F_6 = \frac{900 + 2(700) + 4(600)}{7} = 671.43$$

8.4.3 Regression Analysis

Regression analysis is used to develop a functional relationship between the independent variable being forecasted and one or more independent predictor variables. In time-series regression models, the only independent variable is time. In causal regression models, other independent predictor variables are present. For example if the cost of maintenance for the current period $m(t)$ is a linear function of the number of operational hours in the same period $h(t)$, then the model is given by

$$m(t) = a + bh(t) + \varepsilon_t \quad (8.7)$$

Equation 8.7 represents a straight-line regression relationship with a single independent predictor variable, namely $h(t)$. The parameters a and b are respectively called the intercept and the slope of this line. Regression analysis is the process of estimating these parameters using the least-squares method. This method finds the best values of a and b that minimize the sum of the squared vertical distances (errors) from the line.

The general straight-line equation showing a linear trend of maintenance work demand D_i over time is

$$D_i = a + bt_i + \varepsilon_i \quad (8.8)$$

where

$$D_i = \text{demand at time period } t_i,$$

$$\varepsilon_i = \text{error at time period } t_i$$

Let us assume that n historical data points are available: $(t_1, D_1), (t_2, D_2), \dots, (t_n, D_n)$. The least-squares method estimates a and b by minimizing the following sum of squared errors:

$$SSE = \sum_{i=1}^n (D_i - a - bt_i)^2 \quad (8.9)$$

Taking partial derivatives with respect to a and b and setting them equal to zero produces a 2×2 system of linear equations, whose solution is given by

$$b = \frac{n \sum_{i=1}^n t_i D_i - \sum_{i=1}^n t_i \sum_{i=1}^n D_i}{n \sum_{i=1}^n t_i^2 - \left(\sum_{i=1}^n t_i \right)^2} \quad (8.10)$$

$$a = \frac{1}{n} \left(\sum_{i=1}^n D_i - b \sum_{i=1}^n t_i \right) = \bar{D} - b\bar{t} \quad (8.11)$$

Quite often, the variable being forecasted is a function of several predictor variables. For example, maintenance cost might be a linear function of operating hours, $h(t)$, and the age of the plant, t , which can be expressed as

$$m(t) = a + bh(t) + ct + \varepsilon_i$$

Least-squares regression methodology can easily accommodate multiple variables and also polynomial or nonlinear functional relationships.

Example 8.3: Demand for a given spare part is given below for the last 4 years. Use linear regression to determine the best-fit straight line and to forecast spare part demand in year 5.

Year t	1	2	3	4
Spare part demand $D(t)$	100	120	150	170

Intermediate calculations for the summations needed in Equations 8.10 and 8.11 are shown in Table 8.1 below.

Table 8.1. Data and intermediate calculations for the linear regression example

					Sum
t	1	2	3	4	10
$D(t)$	100	120	150	170	540
$tD(t)$	100	240	450	680	1470
t^2	1	4	9	16	30

Using Equations 8.10 and 8.11, the slope and intercept of the line are estimated as follows:

$$b = \frac{4(1470) - 10(540)}{4(30) - 10^2} = 24, \quad a = \frac{1}{4}[540 - 24(10)] = 75$$

The equation of the least-squares straight line is $D(t) = 75 + 24t$. Therefore, the forecasted spare part demand in year 5 is

$$D(5) = 75 + 24(5) = 195 \text{ units}$$

8.4.4 Exponential Smoothing

8.4.4.1 Simple Exponential Smoothing (ES)

Simple exponential smoothing (ES) is similar to weighted moving average (WMA) in assigning higher weights to more recent data, but it differs in two important aspects. First, WMA is a weighted average of only the last N data points, while ES is a weighted average of *all* past data. Second, the weights in WMA are mostly arbitrary, while the weights in ES are well structured. In fact, the weights in ES decrease exponentially with the age of the data. On the other hand, exponential smoothing is very easy to use, and very easy to update by including new data as it becomes available. In addition, we must save the last N observations for WMA, but need to save only the last observation and the last forecast for ES. These characteristics have made exponential smoothing very popular. Basically, the current forecast is a weighted average of the last forecast and the last actual observation. Given the value of smoothing constant α ($0 \leq \alpha \leq 1$), which is the relative weight of the last observation, the forecast is obtained by

$$F_t = \alpha D_{t-1} + (1 - \alpha)F_{t-1} \quad (8.12)$$

The greater the value of α , the more weight of the last observation, *i.e.*, the quicker the reaction to changes in data. However, large values of α lead to highly variable, less stable, forecasts. For forecast stability, a value of α between 0.1 and 0.3 is usually recommended for smooth planning. The best value of α can be determined from experience or by trial and error (choosing the value with minimum error). It can be shown that the ES forecast is a weighted average of all past data, where the weights decrease exponentially with the age of the data as expressed by

$$F_t = \sum_{i=0}^{\infty} \alpha(1 - \alpha)^i D_{t-i-1} \quad (8.13)$$

Using Equation 8.12, the first forecast F_1 requires the non-existent values of D_0 and F_0 . Therefore, an initial value of F_1 must be specified for starting the process. Usually, F_1 is set equal to the actual demand in the first period D_1 , or to the average of the first few observations.

If simple exponential smoothing at Equation 8.12 is used with linear data ($D_t = a + bt$), then the error will depend on the slope b and the smoothing constant α . As $t \rightarrow \infty$, the forecast will lag behind the actual demand by

$$\lim_{t \rightarrow \infty} \{\varepsilon_t = D_t - F_t\} = \frac{b}{\alpha} \quad (8.14)$$

To make $MA(N)$ and $ES(\alpha)$ consistent, we equate the two lags, ensuring that the distribution of forecast errors will be the same, although individual forecasts will not be the same. Equating the exponential smoothing lag of Equation 8.14 with the moving average lag of Equation 8.4, we obtain the following value for α :

$$\alpha = \frac{2}{N+1} \quad (8.15)$$

Example 8.4: Given that $\alpha = 0.2$ and $F_1 = D_1$, apply simple exponential smoothing to the data of Example 8.1 to forecast maintenance workload in month 6.

Using Equation 8.12, the calculations are shown in Figure 8.2. The forecast for month 6 is $F_6 = 736.32$ man-hours.

Table 8.2. Data and intermediate calculations for the simple exponential smoothing example

t	1	2	3	4	5	6
D_t	800	600	900	700	600	
F_t	800	0.2(800) + 0.8(800) = 800	0.2(600) + 0.8(800) = 760	0.2(900) + 0.8(760) = 788	0.2(700) + 0.8(788) = 770.4	0.2(600) + 0.8(770.4) = 736.32

8.4.4.2 Double Exponential Smoothing (Holt's Method)

The simple exponential smoothing at Equation 8.12 can be used to estimate the parameters for a constant (stationary) model. However, double or triple exponential smoothing approaches can be used to deal with linear, polynomial, and even seasonal forecasting models. Several double exponential smoothing techniques have been developed for forecasting with linear data. One of these is Holt's double exponential smoothing method, which is described below.

Holt's double exponential smoothing method requires two smoothing constants: α and β ($\beta \leq \alpha$). Two smoothing equations are applied: one for a_t , the intercept at time t , and another for b_t , the slope at time t :

$$a_t = \alpha D_t + (1 - \alpha)(a_{t-1} + b_{t-1}) \quad (8.16)$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \quad (8.17)$$

The initial values b_0 and a_0 are obtained as follows:

$$b_0 = \frac{D_n - D_1}{t_n - t_1} \quad (8.18)$$

$$a_0 = \frac{1}{n} \left(\sum_{i=1}^n D_i - b_0 \sum_{i=1}^n t_i \right) = \bar{D} - b_0 \bar{t} \quad (8.19)$$

At the end of period t , the forecast for period τ ($\tau > t$) is obtained as follows:

$$F_\tau = a_t + b_t(\tau - t) \quad (8.20)$$

Example 8.5: Given that $\alpha = \beta = 0.2$, apply Holt's double exponential smoothing method to the data of Example 8.3 in order to forecast spare part demand in year 5.

First, initial conditions are calculated by Equations 8.18 and 8.19:

$$b_0 = \frac{190 - 100}{4 - 1} = 30, \quad a_0 = \frac{1}{4} [570 - 30(10)] = 67.5$$

Intermediate calculations are shown in Table 8.3.

Table 8.3. Data and calculations for the double exponential smoothing example

t	0	1	2	3	4
D_t		100	120	160	190
a_t	67.5	$0.2(100) + 0.8(67.5 + 30) = 98$	$0.2(120) + 0.8(98 + 30.1) = 126.48$	157.005	187.544
b_t	30	$0.2(98 - 67.5) + 0.8(30) = 30.1$	$0.2(126.48 - 98) + 0.8(30.1) = 29.776$	29.926	30.049

The forecasting model at the end of year 4 is: $F_t = 187.544 + 30.049(t - 4)$. Therefore, the forecasted spare part demand in year 5 is given by: $F_5 = 187.544 + 30.049(5 - 4) = 217.593$.

8.4.5 Seasonal Forecasting

Demand for many products and services follows a seasonal or cyclic pattern, which repeats itself every N periods. Although the term “seasonal” is usually associated with the four seasons of the year, the length of the seasonal cycle N depends on the nature of demand for the particular product or service. For example, demand for electricity has a daily cycle, demand for restaurants has a weekly cycle, while demand for clothes has a yearly cycle. The demand for many products may have several interacting cyclic patterns. For example, electricity consumption has daily weekly, and yearly seasonal patterns.

Maintenance workload may show seasonal variation due to periodic changes in demand, weather, or operational conditions. If demand for products is seasonal, then greater production rates during the high-season intensify equipment utilization and increase the probability of failure. If demand is not seasonal, high temperatures during summer months may cause overheating and more frequent equipment failures. Plotting the data is important to judge whether or not it has seasonality,

trend, or both patterns. Methods are presented below for forecasting with stationary seasonal data and seasonal data that has a trend.

8.4.5.1 Forecasting for Stationary Seasonal Data

The model representing this data is similar to the model presented in Equation 8.1, but it allows for seasonal variations:

$$D_t = c_t \mu + \varepsilon_t \quad (8.21)$$

where

c_t = seasonal factor (multiplier) for time period t , $1 \leq t \leq N$,

$$\sum_{t=1}^N c_t = N$$

Given data for at least two cycles ($2N$), four simple steps are used to obtain forecasts for each period in the cycle:

1. Calculate the overall average μ ;
2. Divide each point by the average μ to obtain seasonal factor estimate;
3. Calculate seasonal factors c_t by averaging all factors for similar periods; and
4. Forecast by multiplying μ with the corresponding c_t for the given period.

Example 8.6: The quarterly totals of maintenance work orders are given below for the last 3 years. Forecast the number of maintenance work orders required per quarter in year 4.

	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Year 1	7,000	3,500	3,000	5,000
Year 2	6,000	4,000	2,500	5,500
Year 3	6,500	4,500	2,000	4,500

Step 1 - sum of all data = 54,000

Overall average $\mu = 54,000/12 = 4,500$

Step 2 and 3 - dividing data by 4,500 and averaging columns gives the values in Table 8.4.

Table 8.4. Calculations for the stationary seasonal forecasting example

	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Year 1	1.5556	0.7778	0.6667	1.1111
Year 2	1.3333	0.8889	0.5556	1.2222
Year 3	1.4444	1	0.4444	1
Average = c_t	1.4444	0.8889	0.5556	1.1111

Note that sum of the four seasonal factors ($4444 + \dots + 1.111$) is equal to 4, which is the length of the cycle N (four quarters).

Step 4- finally, the forecasted maintenance work orders for each quarter in year 4 are given by

$$F_1 = 1.4444(4,500) \cong 6,500 \quad \text{Quarter 1}$$

$$F_2 = 0.8889(4,500) \cong 4,000 \quad \text{Quarter 2}$$

$$F_3 = 0.5556(4,500) \cong 2,500 \quad \text{Quarter 3}$$

$$F_4 = 1.1111(4,500) \cong 5,000 \quad \text{Quarter 4}$$

Of course, we could have obtained the forecasts directly by averaging the original data for each shift. However, going through all four steps ensures that the model is completely specified in terms of the mean value μ and seasonal factors $c_1, \dots, c_N$.

8.4.5.2 Forecasting for Seasonal Data with a Trend

It is possible for a time series to have both seasonal and trend components. For example, the demand for airline travel increases during summer, but it also keeps growing every year. The model representing such data is given by

$$D_t = c_t(a + bt) + \varepsilon_t \quad (8.22)$$

The usual approach to forecast with seasonal-trend data is to estimate each component by trying to remove the effect of the other one. Thus, several forecasting methods have been developed for this type of data, all of which basically use the same general approach which is to: (1) remove trend to estimate seasonality, (2) remove seasonality to estimate trend, and (3) forecast using both seasonality and trend. Among the simplest of these methods is the cycle average method, whose steps are described below:

1. Divide each cycle by its corresponding cycle average to remove trend.
2. Average the de-trended values for similar periods to determine seasonal factors $c_1, \dots, c_N$. If $\sum c_t \neq N$, normalize seasonal factors by multiplying them with $N/\sum c_t$.
3. Use any appropriate trend-based method to forecast cycle averages.
4. Forecast by multiplying the trend-based cycle average by appropriate seasonal factor.

Example 8.7: For a university maintenance department, the number of work orders per academic term is given below for the last 3 years. Forecast the number of maintenance work orders required per term in year 4.

	Term 1	Term 2	Term 3 (summer)
Year 1	10,000	7,000	5,000
Year 2	12,000	8,000	6,000
Year 3	14,000	9,000	7,000

Unlike the previous example, the above seasonal data has an increasing trend from year to year. Calculations for seasonal factors (steps 1 and 2) are shown in Tables 8.5 and 8.6.

Table 8.5. Calculating cycle averages

Term: t Year: d	1	2	3	Cycle (year) Sum	Cycle average: A_d
1	10,000	7,000	5,000	22,000	7,333.33
2	12,000	8,000	6,000	26,000	8,666.67
3	14,000	9,000	7,000	30,000	10,000

Table 8.6. Calculating seasonal factors by dividing by cycle averages

Term: t Year: d	1	2	3
1	1.364	0.955	0.682
2	1.385	0.923	0.692
3	1.400	0.900	0.700
Average = c_t	1.383	0.926	0.691

There is no need to normalize seasonal factors since their sum ($383 + 0.926 + 0.691$) is equal to 3, which is the length of the cycle N (three terms).

Using regression, calculations for the trend components of cycle averages (step 3) are shown in Table 8.7.

Table 8.7. Calculating seasonal factors

	d	A_d	dA_d	d^2
	1	7,333.33	7,333.33	1
	2	8,666.67	17,333.33	4
	3	10,000	30,000	9
Sum	6	26,000	54,666.67	14

Using Equations 8.10 and 8.11, the slope and intercept of the cycle averages are estimated as follows:

$$b = \frac{3(54,666.67) - 6(26,000)}{3(14) - 6^2} = 1,333.33$$

$$a = \frac{1}{3} [26,000 - 1,333.33(6)] \cong 6,000$$

The forecasting model for period (term) t of cycle (year) d is given by

$$F_{d,t} = c_t[6,000 + 1,333.33d]$$

Forecasted maintenance work orders required per term in year 4 are calculated as

$$\begin{aligned} F_{4,1} &= 1.383[6,000 + 1,333.33(4)] &= 15,674 && \text{Term 1} \\ F_{4,2} &= 0.926[6,000 + 1,333.33(4)] &= 10,495 && \text{Term 2} \\ F_{4,3} &= 0.691[6,000 + 1,333.33(4)] &= 7,831 && \text{Term 3} \end{aligned}$$

8.4.6 Box-Jenkins Time Series Models

Using the data correlation structure, Box-Jenkins models can provide excellent forecasts, but they require extensive data and complex computations, making them unsuitable for manual calculations. Although autocorrelation analysis is used to find best forecasting model for a given data, judgment plays a role, and the model is not flexible to changes in the data. The two basic Box-Jenkins types are the autoregressive (AR) and the moving average (MA) models. In autoregressive (AR) models, the current value of the time series depends on (is correlated with) previous values of same series. An autoregressive model of order p , which is denoted by $AR(p)$, is given by

$$x_t = a + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (8.23)$$

where

$$\begin{aligned} a, \phi_1 \dots \phi_p &= \text{parameters of fit} \\ \varepsilon_t &= \text{random error} \end{aligned}$$

In moving average (MA) models, the current value of the time series depends on previous errors. In a certain class of problems, the time series x_t can be represented by a linear combination of independent random errors $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ that are drawn from a probability distribution with mean 0 and variance σ_ε^2 . Usually the errors are assumed to be normal random variables. A moving average model of order q , denoted by $MA(q)$, is expressed as

$$x_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_q \varepsilon_{t-q} \quad (8.24)$$

where

$$\begin{aligned} \mu &= \text{mean of the series} \\ \psi_1, \dots, \psi_q &= \text{parameters of fit} \end{aligned}$$

The two models can be combined to form an autoregressive and moving average model of order p and q , denoted by $ARMA(p, q)$. The order of the model is first determined by autocorrelation analysis, and then the values of the parameters are calculated. The aim is usually to find the model that adequately fits the data with the minimum number of parameters. Box and Jenkins (1970) in their book suggested a general methodology for developing an $ARMA(p, q)$ model. The methodology consists of the three following major steps: (1) a tentative model of the $ARMA(p, q)$ class is identified through autocorrelation analysis of the historical data, (2) the unknown parameter of the model are estimated, and (3) diagnostic checks are performed to establish the adequacy of the model or look for potential improvements.

Frequently, several forecasting models could be used to forecast the future maintenance workload. The forecasting techniques presented in the preceding sections may fit the given data with varying degrees of accuracy. In the following section, error analysis is presented as a tool for evaluating and comparing forecasts.

8.5 Error Analysis

If a single forecasting model is applied, error analysis is used to evaluate its performance and to check how closely it fits the given actual data. If several forecasting models are available for a particular set of data, then error analysis is used to compare objectively and systematically the alternative models in order to choose the best one.

The forecasting error ε_t in time period t is defined as the difference between the actual and the forecasted value for the same period:

$$\varepsilon_t = D_t - F_t \quad (8.25)$$

The following error measures are available for checking an individual forecasting model adequacy and comparing among several forecasting models:

1. Sum of the errors (*SOE*)

$$SOE = \sum_{t=1}^n (D_t - F_t) \quad (8.26)$$

Usually used as a secondary measure, *SOE* can be deceiving as large positive errors may cancel out with large negative errors. However, this measure is good for checking bias, *i.e.*, tendency of forecast values to overestimate or underestimate actual values consistently. If the forecast is unbiased, *SOE* should be close to zero.

2. Mean absolute deviation (*MAD*)

$$MAD = \frac{1}{n} \sum_{t=1}^n |D_t - F_t| \quad (8.27)$$

This measure neutralizes the opposite signs of errors by taking their absolute values. If errors are normally distributed, then $1.25 \times MAD$ is approximately equal to the standard deviation of errors, σ .

3. Mean squared error (*MSE*)

$$MSE = \frac{1}{n} \sum_{t=1}^n (D_t - F_t)^2 \quad (8.28)$$

This measure neutralizes the opposite signs of errors by squaring them. If errors are normally distributed, then *MSE* is approximately equal to the variance of errors σ^2 .

4. Mean absolute percent error (*MAPE*)

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{D_t - F_t}{D_t} \right| \quad (8.29)$$

This measure is an independent yardstick for evaluating the “goodness” of an individual forecast. All the other measures only compare different forecasting models relative to each other.

Example 8.8: Given the actual forecasted values in the table below, calculate the different error measure for the associated forecasting model.

t	1	2	3	4	5
D_t	7	10	9	11	14
F_t	6	8	10	12	14

Intermediate calculations are shown in Table 8.8.

Table 8.8. Calculating error measures.

t	1	2	3	4	5	Sum
ε_t	1	2	-1	-1	0	1
$ \varepsilon_t $	1	2	1	1	0	5
ε_t^2	1	4	1	1	0	7
$100 \varepsilon_t/D_t $	14.29	20	11.11	9.09	0	54.49

Therefore

$$\begin{aligned}
 SOE &= 1 \\
 MAD &= 5/5 = 1.0 \\
 MSE &= 7/5 = 1.4 \\
 MAPE &= 54.49/5 = 10.9\%
 \end{aligned}$$

8.6 Forecasting Maintenance Workload

Different types of maintenance workload require different forecasting approaches. Kelly (2006) categorizes maintenance workload into the following types:

1. *First-line maintenance workload:* maintenance jobs are started in the same shift in which problems arise and completed in less than 24.
 - a) *Corrective emergency:* unplanned and unexpected failures that require immediate attention for safety or economic reasons. The frequency of occurrence and the volume of work are random variables, but the volume of maintenance work is usually huge.
 - b) *Corrective deferred minor:* similar to emergency workload, the frequency and volume of maintenance work are random. However, there is no urge for immediate attention. Therefore, maintenance jobs in this category can be delayed and scheduled when the time and conditions are more convenient.
 - c) *Preventive routine:* frequent, short-duration planned maintenance workload, such as inspection, lubrication, and minor part replacement.

2. *Second-line maintenance workload*: maintenance jobs last less than 2 days and require one or few maintenance workers.
 - a) *Corrective deferred major*: very similar to corrective deferred minor maintenance workload, but requires longer times and greater resources.
 - b) *Preventive services*: similar to preventive routine maintenance workload, but the frequency is lower, and the work is usually done offline, usually in the weekend breaks or during scheduled shutdowns.
 - c) *Corrective reconditioning and fabrication*: similar to deferred major maintenance workload, but the work is performed away from the plant, by another group of maintenance workers.
3. *Third-line maintenance workload*: maintenance jobs require maximum demand for resources, long durations and all craft types, at intermediate and long-term intervals.
 - a) Preventive major work (overhauls, *etc.*): less frequent, off-line major preventive maintenance that involves overhauling major pieces of equipment or plant sections.
 - b) Modifications: infrequent, off-line major preventive work that involves process or equipment redesign. This category typically involves the largest capital cost.

Kelly (2006) suggests the following techniques for forecasting the three types of line maintenance workload:

1. *First-line maintenance workload*: a queuing model should be used to represent the size of the first-line maintenance workload. The average maintenance workload is estimated by the average number of man-hours per hour or per day.
2. *Second-line maintenance workload*: the average maintenance workload is estimated by the average number of man-hours per week. This average should be prioritized and updated according to the plant condition.
3. *Third-line maintenance workload*: long-range (5-year) overhaul and shutdown plans are used to predict maintenance workloads and associated resource requirements.

The above discussion focuses on forecasting maintenance workload for existing plants. For new plants, forecasting the maintenance load is more challenging due to the lack of historical data. In such cases, we must revert to qualitative or subjective forecasting techniques presented in Section 8.3.

8.7 Maintenance Capacity Planning

Capacity is the maximum output that can be provided in a specified time period. In other words, capacity is not the absolute volume of work performed or units produced, but the rate of output per time unit. Maintenance capacity planning aims to find the optimum balance between two kinds of capacity: available capacity, and required capacity. Available capacity is mostly constant because it depends on fixed maintenance resources such as maintenance equipment and manpower. On the other hand, required capacity (or maintenance workload) is mostly fluctuating from one period to another according to trend or seasonal patterns.

Effective maintenance capacity planning depends on the availability of the right level of maintenance resources. Resource planning is the process of determining the right level of resources over a long-term planning horizon. Usually, resource planning is done by summing up quarterly or annual maintenance reports and converting them into gross measures of maintenance capacity. Resource planning is a critical strategic function, with serious consequences for errors. If the level of resources is too high, then large sums of capital will be wasted on unused resources. If the level of resources is too little, then lack of effective maintenance resources will reduce the productivity and shorten the life of manufacturing equipment.

Maintenance capacity planning is one function of maintenance capacity management. The other function is maintenance capacity control, in which actual and planned maintenance outputs are compared, and corrective action is taken if necessary. Usually, both available and required capacities are measured in terms of standard work hours. The required capacity for a given period is the sum of standard hours of all work orders, including setup and tooling times. The process of maintenance capacity planning can be briefly described as follows:

1. Estimate (forecast) the total required maintenance capacity (maintenance workload) for each time period;
2. For each time period, determine the available maintenance capacity of each maintenance resource (*e.g.*, employees, contract workers, regular time, and overtime); and
3. Determine the level of each maintenance resource to assign to each period in order to satisfy the required maintenance workload.

The main problem in maintenance capacity planning is how to satisfy the required maintenance workload in each period. Typically, in certain time periods, excessive workload or shortage of available resources necessitate the delay of some work orders to later periods. Therefore, maintenance capacity planning has to answer two questions in order to satisfy the demand for any given period: (1) how much of each type of available maintenance capacity (resource) should be used, and (2) when should each type of resource be used. The usual objective of maintenance capacity planning is to minimize the total cost of labor, subcontracting, and delay (backlogging). Other objectives include the maximization of profit, availability, reliability, or customer service.

Capacity planning techniques are generally characterized by 12-month planning horizons, monthly time periods, fluctuating demand, and fixed capacity. Four basic strategies are used to match the fixed capacity with fluctuating monthly demands:

1. Chase strategy: performing the exact amount of maintenance workload required for each month, without advancing or delay;
2. Leveling strategy: the peaks of demand are distributed to periods of lower demand, aiming to have a constant level of monthly maintenance activity;
3. Demand management: the maintenance demand itself is leveled by distributing preventive maintenance equally among all periods; and
4. Subcontracting: regular employees perform a constant level of monthly maintenance activity, leaving any excess workload to contractors.

The above capacity planning strategies are considered pure or extreme strategies that usually perform poorly. The best strategy is generally a hybrid strategy, which can be found by several available techniques. Capacity planning techniques are generally classified into two main types: deterministic and stochastic techniques. Deterministic techniques assume that the maintenance workload and all other significant parameters are known constants. Two deterministic techniques will be presented in the following section:

1. Modified transportation tableau method; and
2. Mathematical programming.

Stochastic capacity planning techniques assume that the maintenance workload and possibly available capacity and other relevant parameters are random variables. Statistical distribution-fitting techniques are used to identify the probability distributions that best describe these random variables. Since uncertainty always exists, statistical techniques are more representative of real life. However, statistical models are generally more difficult to construct and solve. The two following stochastic techniques will be presented in Section 8.9:

1. Queuing models; and
2. Stochastic simulation.

8.8 Deterministic Approaches for Capacity Planning

The modified transportation tableau method and mathematical programming are presented in the following subsections.

8.8.1 Modified Transportation Tableau Method

For each maintenance craft, the required capacity is given by the forecasted workload for each period. The available capacity for each period is given by the quantity of available resources of different categories. Each of these categories,

such as regular time, overtime, and subcontract, has its own cost. Generally, it is possible to advance some required preventive maintenance work to earlier periods or delay some maintenance work to later periods. However, any advance or delay (backlogging) has an associated cost which is proportional to the volume of shifted work and the length of the time shift. Therefore, the heuristic solution tries to find the least-cost assignment of the required workload in terms of quantity (to different resources) and timing (to different time periods).

The maintenance capacity planning problem is formulated as a transportation model, where the “movement” is not in the space domain, but in the time domain. Maintenance work is “transported” from periods in which the work is performed (sources) to periods where the work is required (destinations). Specifically, each work period is divided into a number of sources that represent the number of maintenance work resources available in the period. Thus, if the planning horizon covers N periods, and if m maintenance resources (e.g., regular, overtime, and subcontract) are available in each period, then the total number of sources is mN . The supply for each source is equal to the capacity of each resource in the given period. The demand for each destination is the required workload for the given period. Notation used in the transportation tableau is defined as:

- c_m = cost of maintenance with resource m per man-hour
- c_A = cost of advancing (early maintenance) per man-hour per unit time
- c_B = cost of backordering (late maintenance) per man-hour per unit time
- $Q_{m,t}$ = capacity of maintenance resource m in period t
- D_t = maintenance demand (required workload) in period t

The total cost of performing maintenance with resource (m) in month (i) to satisfy demand in month (j) is given by

$$TC_{i,j,m} = \begin{cases} c_m + (j-i)c_A, & i \leq j \\ c_m + (i-j)c_B, & i > j \end{cases} \quad (8.30)$$

The transportation tableau in Table 8.9 shows the setup for a three-period planning horizon, with three resources for maintenance work in each period ($m = R$: regular time, $m = O$: overtime, $m = S$: subcontract). Assigning an infinite cost (∞) prohibits assigning any maintenance work to the given (i, j) cell. For example, assigning a cost of (∞) to cells where ($i < j$) would prevent early execution of preventive maintenance work, i.e., execution before the due date.

After the modified transportation tableau is constructed, it is solved by the least-cost assignment heuristic. This heuristic assigns as much as possible (the minimum of supply and demand) to the available (unassigned) cell with the least cost. After each assignment, the supply and demand for the given cell are updated, and the process continues until all demands have been assigned. Although this technique does not guarantee an optimum solution, it is an effective heuristic that frequently leads to optimum solutions.

Table 8.9. Transportation tableau for three-periods and three maintenance resources

Execution Periods	Resource used	Demand periods			Capacity
		1	2	3	
1	Regular time	c_R	$c_R + c_A$	$c_R + 2c_A$	$Q_{R,1}$
	Overtime	c_O	$c_O + c_A$	$c_O + 2c_A$	$Q_{O,1}$
	Subcontract	c_S	$c_S + c_A$	$c_S + 2c_A$	$Q_{S,1}$
2	Regular time	$c_R + c_B$	c_R	$c_R + c_A$	$Q_{R,2}$
	Overtime	$c_O + c_B$	c_O	$c_O + c_A$	$Q_{O,2}$
	Subcontract	$c_S + c_B$	c_S	$c_S + c_A$	$Q_{S,2}$
3	Regular time	$c_R + 2c_B$	$c_R + c_B$	c_R	$Q_{R,3}$
	Overtime	$c_O + 2c_B$	$c_O + c_B$	c_O	$Q_{O,3}$
	Subcontract	$c_S + 2c_B$	$c_S + c_B$	c_S	$Q_{S,3}$
Maintenance demand		D_1	D_2	D_3	

Example 8.9: The required maintenance workload for the next four months is 400, 60, 300, and 500 man-hours, respectively. The demand can be met by either regular time at a cost of \$13 per hour or overtime at a cost of \$20 per hour. Regular time and overtime capacities are respectively 400 and 100 h per month. Early maintenance costs \$3 per hour per month, while late maintenance costs \$5 per hour per month. Using the modified transportation method, develop the capacity plan to satisfy the required workload.

Table 8.10 shows the modified transportation tableau for this example, with hourly costs at the corners of relevant cells. Using the least-cost assignment heuristic, the capacity plan solution is shown in the table, where the highlighted cells indicate active maintenance assignments. The demands of both months 1 and 3 are entirely met by regular time maintenance in the same month. The demand of month 2 is met by regular time maintenance in month 2 in addition to overtime maintenance in months 2 and 3. Finally, the demand of month 4 is met by both regular time and overtime maintenance in month 4. The total cost (TC) of the plan is obtained by multiplying the assigned hours by the corresponding costs:

$$\begin{aligned}
 TC &= 13(400) + 13(400) + 20(100) + 16(100) \\
 &\quad + 13(300) + 13(400) + 20(100) = \$25,100
 \end{aligned}$$

The transportation tableau method is useful for simple cost functions. More complicated relations and cost structures, *e.g.*, the cost of hiring and firing, require more sophisticated methods such as mathematical programming.

Table 8.10. Data and solution of Example 8.9

Execution months	Resources used	Demand months				Capacity
		1	2	3	4	
1	Regular time	13	18	23	28	400
	Overtime	20	25	30	35	100
2	Regular time	16	13	18	23	400
	Overtime	23	20	25	30	100
3	Regular time	19	16	13	18	400
	Overtime	26	23	20	25	100
4	Regular time	22	19	16	13	400
	Overtime	29	26	23	20	100
Maintenance demand		400	600	300	500	

8.8.2 Mathematical Programming Methods

Taha (2003) provides a thorough discussion of mathematical programming models and solution techniques. Mathematical programming is a class of optimization models and techniques that includes linear, nonlinear, integer, dynamic, and goal programming. In general, a mathematical programming model is composed of decision variables, one or more objective functions, and a set of constraints. The objective function(s) and all constraints are functions of the decision variables and other given parameters. In linear programming (LP), all of these functions are linear functions. The objective function is the target of optimization, such as the maximum profit or minimum cost. The constraints are equations or inequalities representing restrictions or limitations that must be respected, such as limited capacity. The decision variables are values under the control of the decision maker, whose values determine the optimality and feasibility of the solution.

A solution, specified by fixed values of the decision variables, is considered optimal if it gives the best value of the objective function, and is considered feasible if it satisfies all the constraints. Optimum solutions of small models can be the Solver tool in Microsoft Excel. Larger models are solved by specialized optimization software packages such as LINDO and CPLEX. In addition to optimal values of decision variables, LP solutions obtained by these packages include values of slacks and surpluses, dual prices, and sensitivity analysis (ranges of given parameters in which the basic solution remains unchanged).

Many variations of mathematical programming models could be constructed for maintenance capacity planning. Depending on the particular situation, the decision variables, objective function, and constraints must be formulated to match the given needs and limitations. For example, options such as overtime,

subcontracting, hiring and firing, and performing early preventive maintenance may or may not be applicable to a given maintenance capacity planning situation. Similarly, each situation calls for a different objective such as minimum cost or maximum safety, reliability, or availability. Examples of different variations of mathematical programming models for maintenance capacity planning are given by Alfares (1999), Duffuaa *et al.* (1999, pp. 139–144), and Duffuaa (2000).

The mixed integer programming model presented below is only a general-purpose example. Different components of this model could be added, deleted, or modified in order to tailor it to a specific maintenance capacity planning application.

Parameters

- $c_A (c_B)$ = cost of advancing (backlogging) each maintenance hour by one month, *i.e.*, cost of early (late) maintenance
 $c_R(c_O)$ = cost of regular time (overtime) maintenance per hour
 c_S = cost of subcontract maintenance per hour
 $c_H (c_F)$ = cost of hiring (firing) one worker
 $n_{R,t}$ = number of regular time work hours per worker in month t
 $n_{O,t}$ = maximum number of overtime hours per worker in month t
 $N_{S,t}$ = number of subcontract work hours available in month t
 D_t = demand (forecast) in month t

Decision variables (for each month t)

- | | |
|------------------------|-----------------------------|
| W_t = workforce size | R_t = regular time hours |
| O_t = overtime hours | S_t = subcontracted hours |
| A_t = advanced hours | B_t = backordered hours |
| H_t = number hired | F_t = number fired |

Objective function

$$\min TC = \sum_{t=1}^T c_R R_t + c_O O_t + c_S S_t + c_A A_t + c_B B_t + c_H H_t + c_F F_t \quad (8.31)$$

Constraints

$$W_t = W_{t-1} + H_t - F_t, \quad t = 1, \dots, T \quad (8.32)$$

$$A_t - B_t = A_{t-1} - B_{t-1} + R_t + O_t + S_t - D_t, \quad t = 1, \dots, T \quad (8.33)$$

$$R_t = n_{R,t} W_t, \quad t = 1, \dots, T \quad (8.34)$$

$$O_t \leq n_{O,t} W_t, \quad t = 1, \dots, T \quad (8.35)$$

$$S_t \leq N_{S,t}, \quad t = 1, \dots, T \quad (8.36)$$

$$W_t, R_t, O_t, S_t, A_t, B_t, H_t, F_t \geq 0, \quad W_t, H_t, F_t \text{ integer}, t = 1, \dots, T \quad (8.37)$$

The objective function at Equation 8.31 aims to minimize the total cost TC of maintenance for all periods in the planning horizon. Constraints at Equations 8.32 and 8.33 respectively balance the workforce size and the maintenance workload between adjacent periods. Constraints at Equations 8.34 and 8.35 respectively relate regular time and overtime work hours to the number of regular maintenance workers in each period. Constraints at Equation 8.36 ensure that the number of assigned subcontract work hours does not exceed the available limit in each period.

Example 8.10: Maintenance workload for the next five months is 2,500, 1,500, 1800, 2800 and 2200 man-hours. This workload can be met by employees on regular time at a cost of \$10 per hour, employees on overtime at a cost of \$15 per hour, or subcontractors at a cost of \$18 per hour. The initial workforce size is 10 employees. Each employee works for 150 regular time hours and a maximum of 60 overtime work hours per month. Maximum capacity of subcontract workers is 200 h per month. Early maintenance costs \$8 per hour per month, while late maintenance costs \$14 per hour per month. For each employee, hiring cost is \$800 and firing cost is \$1000. Assuming zero starting and ending backlog, model and solve this capacity planning problem using mathematical programming.

The integer programming model is given by

$$\begin{aligned} \min TC = & \sum_{t=1}^T 10R_t + 15O_t + 18S_t + 8A_t + 14B_t + 800H_t + 1000F_t \\ \text{subject to} \\ & W_1 = 10 + H_1 - F_1 \\ & W_t = W_{t-1} + H_t - F_t, \quad t = 2, \dots, 5 \\ & A_1 - B_1 = R_1 + O_1 + S_1 - 2500 \\ & A_2 - B_2 = A_1 - B_1 + R_2 + O_2 + S_2 - 1500 \\ & A_3 - B_3 = A_2 - B_2 + R_3 + O_3 + S_3 - 1800 \\ & A_4 - B_4 = A_3 - B_3 + R_4 + O_4 + S_4 - 2800 \\ & 0 = A_4 - B_4 + R_5 + O_5 + S_5 - 2200 \\ & R_t = 150W_t, \quad t = 1, \dots, 5 \\ & O_t \leq 60W_t, \quad t = 1, \dots, 5 \\ & S_t \leq 200, \quad t = 1, \dots, 5 \end{aligned}$$

The optimum solution of the above model was obtained by the optimization software package LINDO. The minimum total cost TC is \$120,920. Decision variables with non-zero values are shown in Table 8.11.

Table 8.11. Integer programming optimal solution of Example 8.10

Month t	1	2	3	4	5
Workforce size W_t	11	11	12	15	15
Regular time hours R_t	1650	1650	1800	2250	2250
Overtime hours O_t	660	0	0	500	0
Subcontract hours S_t	40	0	0	0	0
Backlogged hours B_t	150	0	0	50	0
Hired employees H_t	1	0	1	3	0

8.9 Stochastic Techniques for Capacity Planning

Stochastic models for capacity planning consider various uncertainties ever present in real-life maintenance systems. Uncertainties in maintenance surround both maintenance workload or demand (*i.e.*, timing and severity of equipment failure) and maintenance capacity (*i.e.*, availability and effectiveness of maintenance resources). Usually, uncertainties are represented by probability distributions with specified values of the means and variances. Stochastic models for maintenance capacity planning include queuing models, simulation models, and stochastic programming. Stochastic programming models are mathematical programming models similar to the deterministic models discussed in the previous section, except that some of their elements are probabilistic. Although these models have been used for maintenance capacity planning (*e.g.*, Duffuaa and Al-Sultan, 1999), they are beyond the scope of this chapter, and thus will not be discussed further. In the remainder of this section, queuing theory models and computer simulation models are presented.

8.9.1 Queuing Models

Queuing models deal with systems in which customers arrive at a service facility, join a queue, wait for service, get service, and finally depart from the facility. Queuing theory is used to determine performance measures of the given system, such as average queue length, average waiting time, and average facility utilization (Taha, 2003). In addition, queuing models can be used for cost optimization by minimizing the sum of the cost of customer waiting and the cost of providing service. In applying queuing theory to maintenance systems, the maintenance jobs or required maintenance tasks are considered as the customers, and maintenance resources such as manpower and equipment are considered as the servers.

Queuing systems differ from each other in terms of several important characteristics. To define clearly the characteristics of the given queuing situation, a standard notation (Taha, 2003) is used in the following format:

$$(a/b/c):(d/e/f)$$

where

- a = customer inter-arrival time distribution
- b = service time (or customer departure) distribution
- c = number of parallel servers
- d = queue discipline, *i.e.*, order or priority of serving customers
- e = maximum number of customers allowed in the system (queue plus service)
- f = size of the total potential customer population

Standard symbols are used to represent individual elements of the above notation (symbols a and b). Arrival and service distributions (symbols a and b) are represented by the symbols M (Markovian or Poisson), D (deterministic or constant), E (Erlang or Gamma), and G (general). The queue discipline (symbol d) is represented by the symbols: $FCFS$ (first come, first served), $LCFS$ (last come, first served), $SIRO$ (service in random order), and GD (general discipline). The

symbol M corresponds to the exponential or Poisson distributions. If the inter-arrival time is exponential, then the number of arrivals during a specific period is Poisson. These complementary distributions have a significant role in queuing theory because they have the Markovian (or forgetfulness) property, which makes them completely random. In order to introduce specific queuing models for maintenance capacity planning, the following notation is defined:

- n = number of customers in the system (queue plus service)
- λ_n = customer arrival rate with n customers in the system
- μ_n = customer departure rate with n customers in the system
- ρ = server utilization = λ_n / μ_n
- p_n = probability of n customers in the system
- L_s = expected number of customers in the system
- L_q = expected number of customers in the queue
- W_s = expected waiting time in the system
- W_q = expected waiting time in the queue

Waiting time and the number of customers are directly related by Little's Law, one of the most fundamental formulas in queuing theory:

$$L_s = \lambda_{\text{eff}} W_s, \quad \text{or} \quad L_q = \lambda_{\text{eff}} W_q \quad (8.38)$$

where

λ_{eff} = effective customer arrival rate at the system

Most queuing models are applicable to maintenance capacity planning. Two of these models are presented below, namely the $(M/M/c):(GD/\infty/\infty)$ system and the $(M/M/R):(GD/k/k)$ system.

8.9.1.1 The $(M/M/c):(GD/\infty/\infty)$ System

This queuing system has Markovian inter-arrival and service times, c parallel servers (repairmen), and general service disciplines. Since there are no limits on the number of customers in the system, then $\lambda = \lambda_{\text{eff}}$. Defining $\rho = \lambda/\mu$, the steady-state performance measures for this system are given by

$$L_q = \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} p_0 \quad (8.39)$$

$$L_s = L_q + \rho \quad (8.40)$$

where

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!(1-\rho/c)} \right\}^{-1}, \quad \frac{\rho}{c} < 1 \quad (8.41)$$

The expected number for waiting time in the queue W_q and expected total time in the system W_s are respectively obtained by dividing L_q and L_s by λ .

The above model can be used in maintenance capacity planning to determine the optimum number of servers c (maintenance workers). In this case, the objective would be to minimize the total cost TC of waiting (*i.e.*, cost of equipment downtime) plus the cost of providing maintenance (*i.e.*, cost of maintenance workers). For example, this objective can be expressed as follows:

$$\min TC(c) = c_M c + c_W L_s(c) \quad (8.42)$$

where

c_M = cost of maintenance workers per employee

c_W = cost of waiting time in the queue

It should be noted that Equation 8.42 is only a typical example of a relevant objective in maintenance capacity planning. Several alternative objective functions are possible; for instance, c could be replaced by μ , while L_s could be replaced by L_q , W_s , or W_q .

Example 8.11: A maintenance department repairs a large number of identical machines. Average time between failures is 2 h and 40 min, and average repair time is 5 h; both are exponentially distributed. The hourly labor cost is \$15 per maintenance employee, while the hourly cost of downtime is \$40 per waiting machine. Use queuing theory to determine the optimum number of maintenance employees.

$$\lambda = 1/2.6667 = 0.375$$

$$\mu = 1/5 = 0.2$$

$$\rho = 0.375/0.2 = 1.875$$

Since $\rho/c = 1.875/c < 1$, then $c > 1.875$, or $c \geq 2$

For $c = 2$, the average number of waiting machines $L_s(2)$ and associated total cost $TC(2)$ are calculated by Equations 8.39–8.42 as follows:

$$p_0(2) = \left\{ \sum_{n=0}^{2-1} \frac{1.875^n}{n!} + \frac{1.875^2}{2!(1-1.875/2)} \right\}^{-1} = \frac{1}{31} = 0.03226$$

$$L_q(2) = \frac{1.875^{2+1}}{(2-1)!(2-1.875)^2} \left(\frac{1}{31} \right) = \frac{421.875}{31} = 13.60887$$

$$L_s(2) = 13.60887 + 1.875 = 15.48387$$

$$TC(2) = 15(2) + 40(15.48387) = 649.35$$

For $c \geq 3$, the average number of waiting machines $L_s(c)$ and associated total cost $TC(c)$ are similarly calculated by Equations 8.39–8.42. Because $TC(c)$ is convex, we should start with $c = 2$ and increment c by one employee at a time until the total cost $TC(c)$ begins to increase. The calculations are summarized in Table 8.12, showing that the optimum number of maintenance employees is equal to 4.

Table 8.12. Queuing model solution of Example 8.11

c	$p_0(c)$	$L_s(c)$	$TC(c)$
2	0.03226	15.48387	649.35
3	0.13223	2.52066	145.83
4	0.14924	2.00265	140.11
5	0.15255	1.90328	151.13

8.9.1.2 The (M/M/R):(GD/K/K) System

This queuing system is called the machine repair or machine servicing model. It has Markovian inter-arrival and service times, R parallel servers (repairmen), and a general service discipline. This model represents the situation in a shop with K machines (customers). Therefore, K is both the maximum number of customers in the system and the size of the customer population. For this system, the number of repairmen must not exceed the number of machines, *i.e.*, $R \leq K$. Assuming that λ is break-down rate per machine, the steady-state results for this system can be derived as follows:

$$L_s = \sum_{n=0}^K n p_n \quad (8.43)$$

$$L_q = \sum_{n=R+1}^K (n-R) p_n \quad (8.44)$$

where

$$p_n = \begin{cases} \binom{K}{n} \rho^n p_0, & 0 \leq n \leq R \\ \binom{K}{n} \frac{n! \rho^n}{R! R^{n-R}} p_0, & R \leq n \leq K \end{cases} \quad (8.45)$$

$$p_0 = \left(\sum_{n=0}^R \binom{K}{n} \rho^n + \sum_{n=R+1}^K \binom{K}{n} \frac{n! \rho^n}{R! R^{n-R}} \right)^{-1} \quad (8.46)$$

The values of W_q and W_s can be calculated by respectively dividing L_q and L_s by λ_{eff} , which is given by

$$\lambda_{\text{eff}} = \lambda(K - L_s) \quad (8.47)$$

Example 8.12: A manufacturing facility has 27 identical machines. On average, each machine fails every 4 h. For each machine failure, average repair time is 30 min. Both the time between failures and the time for repair are exponentially distributed. The hourly cost for each repair station is \$18, while the hourly cost of lost production is \$55 per broken machine. Apply queuing theory to determine the optimum number of repairmen for this facility.

$$\begin{aligned} \lambda &= 1/4 &= 0.25 \\ \mu &= 1/0.5 &= 2 \\ \rho &= 0.25/2 &= 0.125 \\ K &= 27 \end{aligned}$$

Starting with $R = 1$, $p_0(1)$, $p_1(1)$, and $L_s(1)$ are calculated by equation 8.43, 8.45, and 8.46 as follows:

$$p_0(1) = \left(\sum_{n=0}^1 \binom{27}{n} 0.125^n + \sum_{n=1+1}^{27} \binom{27}{n} n! 0.125^n \right)^{-1} = \frac{1}{13,424,835.87}$$

$$p_n(1) = \binom{27}{n} \frac{n! 0.125^n}{13,424,835.87}, \quad n = 1, \dots, 27$$

$$L_s(1) = \sum_{n=0}^{27} n p_n(1) = 19.000000$$

$$TC(1) = 18(1) + 55(19) = 1063.00$$

Since $TC(R)$ is convex, R is incremented by one at a time until $TC(R)$ starts to increase. The calculations are summarized in Table 8.13. The optimum number of repair stations is equal to five.

Table 8.13. Queuing model solution of Example 8.12

R	$L_s(R)$	$TC(R)$
1	19	1063.00
2	11.07248	644.99
3	5.49428	356.19
4	3.67456	274.10
5	3.18612	265.24
6	3.04971	275.73
7	3.01224	1063.00

8.9.2 Stochastic Simulation

Simulation is a technique in which a computer model is constructed of a real-life system. This model allows us to observe the changing behavior of the system over time and to collect information about the required performance measures. In addition, this technique allows us to perform experiments on the simulation model that would be too expensive, too dangerous, or too time-consuming to perform on the real system. These experiments are performed by running the model under different conditions or assumptions (called scenarios) corresponding to different real-life options. Statistical inference techniques are used to analyze and interpret the results of simulation experiments.

According to Banks *et al.* (2005), simulation models are classified as static or dynamic, deterministic or stochastic, and discrete or continuous. A static (or Monte Carlo) model represents a system at a single given point in time. A dynamic model represents a system over a whole range of different time periods, showing the changing behavior of the system over time. A deterministic model is completely certain, because it has no random variables. A stochastic model includes uncertainty in the form of random variables with specific probability distributions. In discrete simulation models, the system variables change discretely at specific points in time. In continuous simulation models, the system variables may change continuously over time.

Banks *et al.* (2005) propose the following 12-step procedure for building a sound simulation model:

1. Problem formulation: develop a clear statement of the problem.
2. Setting of objectives and overall project plan: specify the question to be answered by simulation, the alternative systems (scenarios) to be considered, and criteria to evaluate those alternatives.
3. Model conceptualization: construct a simulation model of the real system, as simple as possible while capturing all the essential elements.
4. Data collection: collect data to run and to validate the simulation model. This step is time consuming and interrelated with model conceptualization.
5. Model translation: program the model in computer simulation software.
6. Model verification: debug the program to ensure the model's logical structure is correctly represented in the computer.
7. Model validation: compare the model to actual system, and calibrate the model to make its performance measures as close possible to those of the actual system.
8. Experimental design: determine length of initialization period, length of simulation runs, and number of replications of each run.
9. production runs and analysis: run the model, collect performance measures, and analyze results.
10. Additional runs: based on analysis, perform more runs if needed;
11. Documentation and reporting: prepare program documentation and manuals, in addition to reporting on simulation results and recommendations.
12. Implementation: apply approved recommendations.

According to Kelly (2007), simulation allows us to consider many complex features of maintenance systems that cannot be easily included otherwise, such as redundant components, stand-by equipment, aging of machines, imperfect repairs, and component repair priorities. Simulation has been effectively used for maintenance capacity planning because it is capable of handling the inherent uncertainty and complexity in maintenance processes. For example, simulation has been used for determining the optimum number and schedule of maintenance workers, the optimum preventive maintenance policy, and the optimum buffer capacity between pairs of successive machines in a production line. Duffuaa *et al.* (1999) consider simulation well suited for maintenance capacity planning, because of the following characteristics of maintenance systems:

- Complex interaction between maintenance functions and other technical and engineering functions.
- High interdependence of different maintenance factors on each other.
- Prevalence of uncertainty in most maintenance processes.

Numerous simulation models have been proposed for different maintenance systems. For example, Sohn and Oh (2004) use simulation to determine the optimal repair capacity at an IT maintenance center. Duffuaa *et al.* (2001) propose a

generic conceptual simulation model that provides a general framework for realistic simulation models of maintenance systems. This model consists of seven modules:

1. Input module: provides all required data for the simulation model.
2. Maintenance load module: generates the maintenance workload.
3. Planning and scheduling module: assigns available resources to maintenance jobs and schedules them to meet workload requirements.
4. Materials and spares module: ensures availability of materials and supply for maintenance jobs.
5. Tools and equipment module: ensures availability of tools and equipment for maintenance jobs.
6. Quality module: ensures the quality of maintenance jobs.
7. Performance measures module: calculates various performance measures of the maintenance system.

Example 8.13: Alfares (2007) presents a simulation model for days-off scheduling of multi-craft maintenance employees. The maintenance workforce of an oil and gas pipelines department is composed of air conditioning (AC), digital (DG), electrical (EL), machinist (MA), and metal (ME) technicians. Using the workdays/off-days notation, maintenance workers can be assigned to only 3 days-off schedules: (1) the 5/2 schedule, (2) the 14/7 schedule, and (3) the 7/3-7/4 schedule. The simulation model considers stochastic workload variability, limited manpower availability, and employee work schedules. A simplified flowchart of the simulation model is shown in Figure 8.2. The model recommended optimum days-off assignments for the multi-craft maintenance workforce. These assignments are expected to reduce the time in the system W_s by an average of 25% for pipeline maintenance work orders.

8.10 Summary

This chapter presented the basic ideas and procedures in maintenance forecasting and capacity planning. Forecasting has been defined as the prediction of future values, which forms the basis for effective planning. Forecasting techniques are

classified into qualitative (subjective) and quantitative (objective). Subjective techniques are used in the absence of reliable numerical data, and include benchmarking, sales force composite, customer surveys, executive opinions, and the Delphi method. Quantitative or objective forecasting techniques are classified into time-series and causal models. Quantitative forecasting techniques presented for stationary, linear, and seasonal data include the moving average, exponential smoothing, least-squares regression, seasonal forecasting and Box-Jenkins ARMA models

Error analysis was presented as an objective tool to evaluate and compare alternative forecasting models. Different forecasting approaches were presented to

deal with the three types of line maintenance workload requirements: first-line, second-line, and third-line maintenance workloads.

Planning has been defined as preparing for the future, and it must be based on forecasting. Maintenance capacity planning aims to best utilize fixed maintenance resources in order to meet the fluctuating maintenance workload. Therefore, it has to determine when and how much of each type of available maintenance resources should be used. Capacity planning techniques are classified into deterministic and stochastic techniques. Deterministic techniques contain parameters that are known constants, and they include the modified transportation tableau method, and mathematical programming. Stochastic techniques contain parameters that are random variables, and they include queuing models and stochastic simulation.

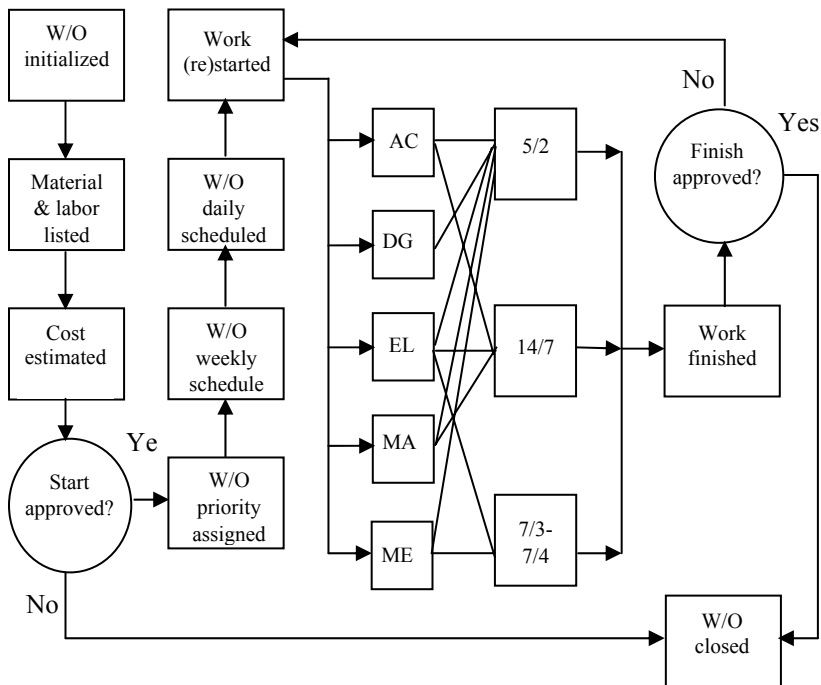


Figure 8.2. Simplified flowchart of the maintenance work order process

References

- Alfares HK (1999) Aircraft maintenance workforce scheduling: a case study. *J of Qual in Maint Eng* 5: 78–88
- Alfares HK (2007) A simulation approach for stochastic employee days-off scheduling. *Int J of Model and Simul* 27: 9–15

- Banks J, Carson JS II, Nelson BL, Nicol DM (2005) Discrete-Event Simulation, 4th Edition, Prentice Hall, Upper Saddle River, USA.
- Box GP, Jenkins GM (1970) Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, USA.
- Duffuaa SO (2000) Mathematical Models in Maintenance planning and scheduling, in Ben-Daya M, Duffuaa SO, Raouf A (eds.), Maintenance Modeling and Optimization, Kluwer, Boston, USA, 39–53
- Duffuaa SO, Al-Sultan KS (1999) A stochastic programming model for scheduling maintenance personnel. Appl Math Model 23: 385–397
- Duffuaa SO, Ben-Daya M, Al-Sultan KS, Andijani AA (2001) A generic conceptual simulation model for maintenance systems. J of Qual in Maint Eng 7: 207–219
- Duffuaa SO, Raouf A, Campbell JD (1999) Planning and Control of Maintenance Systems: Modeling and Analysis, John Wiley & Sons, New York, USA
- Kelly AD (2006) Managing Maintenance Resources, Elsevier
- Kelly AD (2007) The Maintenance Management Framework, Springer, London, 157–184
- Nahmias S (2005) Production and Operations Analysis, 5th Edition, McGraw-Hill, Singapore.
- Rowe, G and Wright, G, (2001) Expert Opinions in Forecasting. Role of the Delphi Technique. In Armstrong (Ed.), Principles of Forecasting: A Handbook of Researchers and Practitioners, Kluwer, Boston, 125–144.
- Sohn SY, Oh KY, (2004) Simulation study for the optimal repair capacity of an IT maintenance center based on LRD failure distribution. Comput & Oper Res 31: 745–759
- Taha H, (2003) Operations Research: An Introduction, 7th Edition, Prentice Hall, Upper Saddle River, USA

Integrated Spare Parts Management

Claver Diallo, Daoud Aït-Kadi, and Anis Chelbi

9.1 Introduction

Maintenance strategies are designed and implemented in order to reduce the frequency and duration of service interruptions, while satisfying constraints on budget, productivity, space, *etc.* A maintenance strategy is defined as the set of actions pertaining to maintaining or restoring a system in a specified state or in a state of readiness to accomplish a certain task. The main scientific contributions dealing with maintenance policies generally address the three following issues in a separate or combined way: the choice and the sequence of actions defining each strategy, the costs and durations of these actions, and the equipment lifetime and repair distributions.

For many companies, the expenses incurred for keeping spare parts until they are used increase significantly the cost of their finished goods. Huge costs related to the inventory management of those parts have triggered studies on the provisioning and management decisions made in the process of acquiring and holding spare parts stocks.

The aim of these spare parts stocks is to protect from long maintenance downtime of randomly failing equipment. This technical maintenance downtime can be severely affected by supply lead-time when replacement parts are not available on-hand. However, the spare part inventory related costs do not permit to keeping spare parts for all failure prone components.

Spare parts are designed for specific usage, their consumption is highly random, and their replenishment lead-times are variable and often unknown. Moreover, these parts can be subject to obsolescence or degradation while in stock, and they are also hardly resalable. Therefore, a procedure is needed to select the components which should have spare parts in stock. The composition of this

package of spare part is established based on technical, economic and strategic considerations. Usually, companies buy their spare parts directly from their original equipment manufacturer (OEM) who is not always easily accessible. The spare parts kit suggested by the OEM is rarely reviewed or questioned, since the clients usually do not have the required knowledge of the equipment to estimate its components' lifetime and reliability. These spare parts kits are generally established based on the knowledge and expertise of the OEM and can be biased by commercial interests (see Dorsch, 1998).

The aim of this chapter is to propose an integrated spare part inventory management approach for multi-components systems subjected to random failures. The content of the paper is structured as follows: Section 9.1 addresses the identification and classification process; Section 9.2 deals with the determination of the spare parts quantity required to achieve pre-determined performance; Section 9.3 is dedicated to the inventory control policies; Section 9.4 tackles the joint-optimization of maintenance and inventory control, optimal reuse policies of used components are addressed in Section 9.5; Section 9.6 deals with the collaborative management of spare parts.

9.2 Spare Parts Identification and Classification

Spare parts identification process is usually initiated from technical considerations. However, the efficiency of this process is affected by the quantity and quality of lifetime information available. At the stage of acquiring equipment, not much information is available to the buyer. Therefore spare parts provisioning decisions are based on the OEM spare parts kits, on failure rates from similar equipment, and on estimates from experts. The initial provisioning problem is addressed by Burton and Jacquette (1973), Geurts and Moonen (1992), and Haneveld and Teunter (1997). For the remainder of the article, it is assumed that a complete set of lifetime data is available either from accelerated tests conducted by the OEM or from long enough operation of the equipment. Hence, for a component i , the lifetime function $f_i(\cdot)$ can be determined according to the methodology depicted in Figure 9.1. Once $f_i(\cdot)$ is known, the reliability function $R_i(\cdot)$, the failure rate $r_i(\cdot)$ and the mean time between failures MTBF can be computed (see Table 9.1). Spare parts are to be acquired if

$$F_i(t) > F^*$$

where F^* is the failure risk the buyer is willing to accept during the mission duration t .

The identification process can also be based on other criteria such as: availability $A(t)$, criticality index from FMECA, importance factor, *etc.*

Once the selection step is applied to all components in the equipment, a list of potential spare parts is obtained. These parts are then ranked through a Pareto classification or multi-criteria method considering technical, economic, and operational criteria (see Braglia *et al.* 2004; Chelbi and Aït-Kadi, 2002; Eisenhower *et al.* 2002; Gajpal *et al.* 1994; Schärliig, 1985). The outcome is a list of components ranked according to their importance to the buyer's production

system, which then leads to the selection of a reduced number of components to make up the final spare parts list. Such a classification is useful whenever the list of potential spare parts is very long and the available resources (storage, budget, personnel) are limited.

Once the spare parts are identified, one has to determine the required quantities to be acquired during a given time period in order to achieve the expected performance levels.

Table 9.1. Basic reliability relations

	$f(t)$	$F(t)$	$R(t)$	$r(t)$
$f(t)$	–	$\int_0^t f(x)dx$	$\int_t^\infty f(x)dx$	$\frac{f(t)}{\int_t^\infty f(x)dx}$
$F(t)$	$\frac{dF(t)}{dt}$	–	$1 - F(t)$	$\frac{dF(t)}{[1 - F(t)]dt}$
$R(t)$	$\frac{-dR(t)}{dt}$	$1 - R(t)$	–	$\frac{-dR(t)}{R(t)dt}$
$r(t)$	$r(t) \cdot e^{-\int_0^t r(x)dx}$	$1 - e^{-\int_0^t r(x)dx}$	$e^{-\int_0^t r(x)dx}$	–

9.3 Determination of the Required Quantity of Spare Parts

For each component on the spare parts list established in Section 9.1, the required quantity to acquire during the equipment economical lifetime must be determined. The following four main procedures will be presented: recommendations from OEM or experts, analytical methods based on reliability or availability, forecasting and simulation.

9.3.1 Recommendations

In case of a lack of useful failure or consumption data, the decision of how many spare parts to buy is based on the recommendations from the OEM, and from surveys of the equipment primary users (operators, mechanics, *etc.*). Consumption records from similar equipment can be used to obtain a significant estimate of the required spare parts number.

9.3.2 Reliability and Availability Based Procedures

When failure records are available, they can be exploited to determine the lifetime density function as shown by Figure 9.1. In such a case, the analytical procedures can be used. The reliability-based procedure is exposed first and followed by the availability-based procedure.

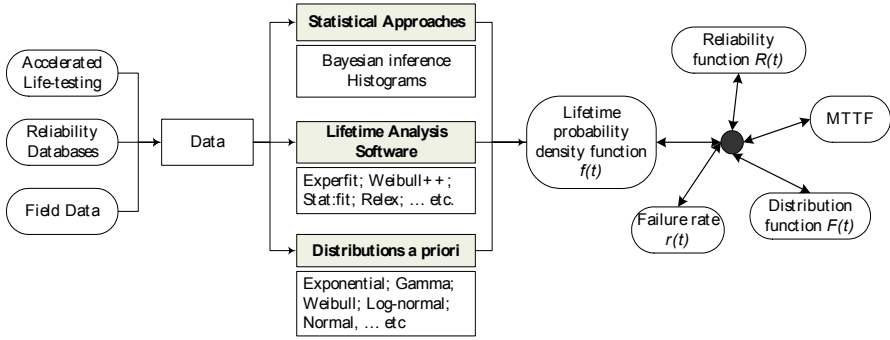


Figure 9.1. Failure data processing diagram

For a component with lifetime density function $f(t)$ and negligible replacement duration, the average number of replacements at failure $M(t)$, with replacements carried-out with new spare parts during a mission of length t , satisfies the following fundamental renewal equation:

$$M(t) = F(t) + \int_0^t M(t-x)f(x)dx$$

If $F^{(i)}(t)$ denotes the i -fold convolution of $F(t)$ with itself, then

$$M(t) = \sum_{i=1}^{\infty} F^{(i)}(t)$$

If at failure the component is minimally repaired without affecting its failure rate $r(t)$, then the average number of failures during the time interval $[0, t]$ is given by

$$M(t) = \int_0^t r(x)dx$$

When the repair or replacement durations are random, the average number of failures during the time interval $[0, t]$ is given by

$$M(t) = \sum_{i=1}^{\infty} G^{(i)}(t)$$

where $G^{(i)}(t)$ denotes the i -fold convolution of $G(t)$ with itself, $g(t)=dG(t)/dt$ being the convolution of the lifetime density function $f(.)$ with the repair or replacement duration density function $h(.)$ such as

$$g(t) = \int_0^t f(t-x)h(x)dx$$

Closed-form expressions for the renewal function $M(t)$ are only known to a relatively short list of distributions used in reliability and maintenance modelling, such as the Uniform, Exponential and Erlang distributions. However, several numerical methods have been proposed to compute $M(t)$ (see Aït-Kadi and Chelbi, 1998; Xie, 1989; Zhang and Jardine, 1998). Diallo and Aït-Kadi (2007) have proposed an approximation based on the Dirac function to compute $g(t)$ and $M(t)$

when repair or replacement durations are not negligible. Once $M(t)$ is known, the average number n of spare parts required for the time interval $[0, t]$ is obtained by rounding it to the next integer:

$$n = \lceil M(t) \rceil.$$

For a component replaced at failure and after T units of time, according to the age replacement policy (ARP), the upper bound of the expected number n_{ARP} of spare parts for a mission of length t is given by Barlow and Proschan (1965):

$$n_{ARP}(t) = \left\lceil \frac{t \cdot [1 - R(T)]}{\int_0^T R(x) dx} \right\rceil$$

If the component is replaced at failure or at predetermined instants kT ($k=1,2,3,\dots$) regardless of its age and state, according to the block replacement policy (BRP), then the expected number n_{BRP} of spare parts for a mission of length t is given by

$$n_{BRP} = \lceil k[M(T) + 1] + M(t - kT) \rceil \text{ with } kT \leq t < (k+1)T$$

For some applications, a spare part can be considered as a stand-by component in the reliability point of view. The determination of the system reliability $R_s(t, n)$ for a stand-by structure with n components allows to calculate the number $n-1$ of spares to keep in stock to achieve a desired reliability level R^* for a given mission duration t . This is equivalent to finding the smallest integer n^* satisfying the following equation:

$$R_s(t, n) \geq R$$

$$\int_t^\infty f^{(n+1)}(x) dx \geq R$$

Where $f^{(i)}(t)$ denotes the i -fold convolution of $f(t)$ with itself.

Once $R_s(t, n)$ is known the number of spare parts to keep can be computed using a simple iterative algorithm from Aït-Kadi *et al.* (2003a).

For a repairable component (as good as new) with failure rate $r(t)=\lambda$ and repair rate $\mu(t)=\mu$, the expression of $R_s(t, n)$ is given by

$$R_s(t, n) = e^{-zt}$$

where

$$z = \frac{(1-\gamma)^2 \gamma^{n^*(t)} \lambda}{1 - \gamma^{[n^*(t)+1]} [1 + [n^*(t) + 1](1-\gamma)]}$$

and

$$\gamma = \lambda/\mu.$$

Table 9.2 from Aït-Kadi *et al.* (2003a) gives the expressions of $R_s(t, n)$ for different configurations of the problem. In the general case, when k components are on operation and $(n-k)$ are kept in stock, the expression of $R_s(t, n)$ becomes

$$R_s(t, n) = e^{-k\lambda t} \sum_{j=0}^{n-k} \frac{(k\lambda t)^j}{j!}$$

Note also that the availability $A_s(t, n)$ could be used instead of the reliability $R_s(t, n)$.

Let us now consider a fleet of N independent and identically distributed (i.i.d.) machines each having a failure rate $\lambda(t)$ and repair rate $\mu(t)$. A stock of y spare machines is held. Whenever one of the N operating machines fails, it is replaced with a spare one. The failed machine is brought to the repair shop equipped with c parallel repair channels. The broken machine is repaired as new and added to the spare machines stock. The process stops as soon as $N+y$ machines are simultaneously broken. It is also known as the repair-man problem (see Gross *et al.* 1977; Sherbrooke, 1968; Taylor and Jackson, 1954). The term “machine” is used in its broader sense: it can designate a complex machinery, a module or a component. Figure 9.2 depicts the main components of the logistical support model required to maintain this fleet of machines in an operating state.

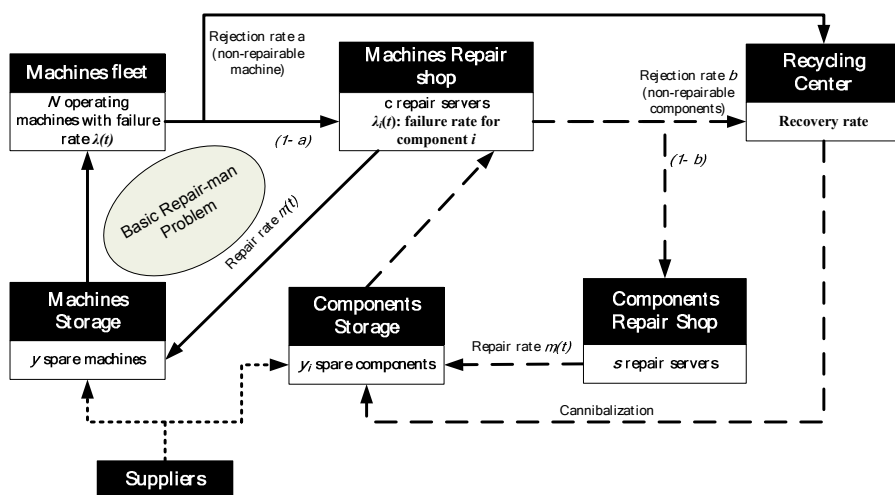


Figure 9.2. Main components of the logistical support model for a fleet of machines

Table 9.2. Reliability expressions for several configurations

λ_i and μ_i : respectively, the failure and repair rates of the i th component, $R_0(t)$: reliability of the component while waiting (in stand-by), $f_0(t)$: pdf of the component lifetime while waiting, $R_1(t)$: reliability of the operating component, $f_1(t)$: pdf of the operating component lifetime.	
Exponential lifetime distribution – independent components – without repair – no failure when in stock	
$\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_n$	$R_s(t, n) = \sum_{i=1}^n \frac{\lambda_1 \lambda_2 \dots \lambda_{i-1} \lambda_{i+1} \dots \lambda_n}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)} \cdot e^{-\lambda_i t}$
Exponential lifetime distribution – i.i.d. components – without repair – no failure when in stock and $\lambda_1 = \dots = \lambda_n = \lambda$	
k operating components, $n-k$ components are waiting	$R_s(t, n, k) = e^{-k\lambda t} \cdot \sum_{i=0}^{n-k} \frac{(k\lambda t)^i}{i!}$
Exponential lifetime distribution – i.i.d. components – with repair – no failure when in stock.	
$\lambda_1 = \dots = \lambda_n = \lambda$	$R_s(t, n) = e^{-z t} \quad \text{where} \quad z = \frac{(1-\gamma)^2 \gamma^{n-1} \lambda}{1-\gamma^n [1+n(1-\gamma)]} \quad \text{with} \quad \gamma = \frac{\lambda}{\mu}$
General lifetime distribution – i.i.d. components – without repair – no failure when in stock.	
k operating components, $n-k$ components are waiting	$R_s(t, n) = [R(t)]^k \sum_{j=0}^{n-k} \frac{[-\ln R(t)]^j}{j!}$
General lifetime distribution – i.i.d. components – without repair – no failure when in stock.	
Three components	$R_s(t) = R(t) + \int_0^t f(\tau) R(t-\tau) d\tau + \int_0^t d\tau_2 \int_0^{\tau_2} f(\tau_1) f(\tau_2) R(t-\tau_2) d\tau_1$ with $0 < \tau_1 < \tau_2 < t$
General lifetime distribution – independent components – without repair – failure when in stock.	
Two components	$R_s(t) = R_1(t) + \int_0^t f_1(\tau) R_0(\tau) R_1(t-\tau) d\tau$ where $0 < \tau < t$

The number y of spares to keep in stock should allow one to reach a given service level NS defined as the probability of having at least one machine in stock. The problem is then to find the smallest $y > 1$ such as

$$\sum_{i=0}^{y-1} P_i \geq NS \quad (9.1)$$

where P_i is the steady-state probability of having i broken machines awaiting repair or being repaired.

When the failure and repair rates are constant such as $\lambda(t)=\lambda$ and $\mu(t)=\mu$, then the P_i are given by Gross *et al.* (1977):

$$P_i = \begin{cases} \left. \begin{array}{l} \frac{N}{i!} \left(\frac{\lambda}{\mu} \right)^i P_0 ; & 0 \leq i < c \\ \frac{N^i}{c^{i-c} c!} \left(\frac{\lambda}{\mu} \right)^i P_0 ; & c \leq i \leq y \\ \frac{N^y N!}{(N-i+y)! c^{i-c} c!} \left(\frac{\lambda}{\mu} \right)^i P_0 ; & y < i \leq y+N \end{array} \right\} \text{ for } c \leq y \\ \left. \begin{array}{l} \frac{N}{i!} \left(\frac{\lambda}{\mu} \right)^i P_0 ; & 0 \leq i < y \\ \frac{N^y N!}{(N-i+y)! i!} \left(\frac{\lambda}{\mu} \right)^i P_0 ; & y \leq i \leq c \\ \frac{N^y N!}{(N-i+y)! c^{i-c} c!} \left(\frac{\lambda}{\mu} \right)^i P_0 ; & c < i \leq y+N \end{array} \right\} \text{ for } c > y \end{cases}$$

The previous result relies on the assumption of exponentiality of failure interarrival and service times [*i.e.*, $\lambda(t)=\lambda$ and $\mu(t)=\mu$]. Often this is not the case in practice. Gross (1976) studied the sensitivity of the model to the exponentiality assumption and derived some rules of thumb for the estimation of the error induced by the assumption.

The models and methods presented so far require at least the knowledge of the lifetime density function for each component. If the available field data lacks the accuracy or the quantity required for the extraction of the lifetime density function, but is sufficient to extract the spare parts consumption or demands at the storehouse, then forecasting techniques can be applied to determine the number of spare parts to provision.

9.3.3 Forecasting Procedure

Forecasting models are widely used to predict the levels of activities in the future based on observations carried-out in the past. Major commercial softwares primarily using quantitative forecasting methods are listed by Yurkiewicz (2006). When spare parts demand forecasts are made, it is necessary to take into account the influence of certain seasonal or temporary factors on the number of breakdowns. Changes in operating conditions (environment, seasons, *etc.*) in production, mechanical loads variations, constitute as many seasonal or transient

factors which affect the failure rate and therefore the number of spare parts used. It is thus necessary to account for those changes when selecting a forecasting model or to envisage an adjustment mechanism (see Figure 9.3).

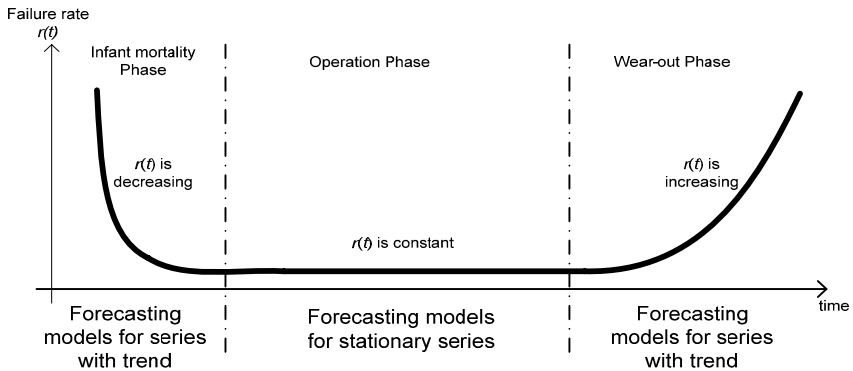


Figure 9.3. Forecasting model selection according to the failure rate profile

In inventory control, a distinction is made between slow-moving and fast moving items. Slow-moving items have an average lead-time demand lower than 10 units per period (see Silver and Peterson 1985). For fast-moving items, the traditional forecasting methods such as exponential smoothing, moving average, and regressions, are effective methods to predict the needs. The slow-moving items can be subdivided into two classes: those with non-intermittent demand and those with intermittent demand. An intermittent demand is a random demand with a great proportion of zero values (Silver' 1981). Methods such as exponential smoothing and moving average are recommended for non-intermittent slow-moving items. Croston (Syntetos and Boylan, 2005; Willemain *et al.* 1994) and bootstrap methods are recommended for intermittent slow-moving items. Spare parts and insurance-type spare parts are generally intermittent slow-moving items. A spare parts forecasting techniques selection guide summary is provided in Table 9.3.

Table 9.3. Spare parts forecasting techniques selection guide

	Demand	Suggested forecasting models
Fast moving items	Stationary (constant failure rate)	Simple moving average Weighted moving average Exponential smoothing
	Stationary (constant failure rate)	Moving average Two parameters exponential smoothing Linear regression
Slow moving items	Non-intermittent	Moving average Exponential smoothing
	Intermittent	Croston method Bootstrap method Moving average

The intermittent demand is also often erratic, *i.e.*, there is a great variability between the non-zero values. Brown (1977) considers that a demand is erratic when its standard deviation is higher than its average. It should be noted that a vast majority of the commercial softwares still use exponential smoothing for intermittent slow-moving items, even if the Croston and bootstrap method are known to be very efficient. The Croston method can easily be implemented in a spreadsheet application, making it very appealing for actual use in real-life applications.

Instead of smoothing the demand at each period, the Croston method applies exponential smoothing to both demand size and inter-demand interval. The smoothing procedures are applied only to the periods with non-zero demand.

Let:

x_n = demand at period n ;

α_i = smoothing constant used for updating the inter-demand interval;

α_d = smoothing constant used for updating the demand size;

$\hat{m}_n$ = estimate of the average interval between consecutive demand incidences;

$\hat{X}_n$ = estimate of the average size of demand;

$\hat{X}_{n,n+t}$ = estimate of the average size of demand per period computed at the end of period n for period $n+t$.

The Croston method as modified by Syntetos and Boylan (2005) gives:

$$\hat{X}_n = \alpha_d x_n + (1 - \alpha_d) \hat{X}_{n^*} \quad 0 < \alpha_d < 1$$

$$\hat{m}_n = \alpha_i (n - n^*) + (1 - \alpha_i) \hat{m}_{n^*} \quad 0 < \alpha_i < 1$$

$$\hat{X}_{n,n+t} = \left(1 - \frac{\alpha_i}{2}\right) \frac{\hat{X}_n}{\hat{m}_n} \quad t = 1, 2, 3, \dots$$

where n^* is the index of the period where the previous smoothing took place (previous period with non-zero demand) and usually $\alpha_i = \alpha_d$.

9.3.4 Simulation

For several cases, the failure and repair processes are so complex and intricate that it is mathematically cumbersome to model and determine the required number of spare parts. In these cases, simulation can be used to model the failure and repair/replacement processes. Simulation is also an excellent alternative when the lifetime and service distributions functions are not exponential. The principle of simulation consists in the random generation of the instants of breakdown and repair duration according to their respective distribution functions. At breakdown, a spare part is taken from the spare parts stock if there is any available. At the end of each repair, a spare part is added to the stock. The service level is the proportion of time that a request for a spare part is filled from the shelves. Each computational reproduction of the behaviour of the system is called replication. By generating a great number of replications, it is possible to obtain an average result similar to the

actual behaviour of the system. See Diallo (2006) and Sarker and Haque (2000) for examples relying on simulation to determine the required quantity of spare parts.

Having determined the components to keep in stock and their quantities for a given period, the next section will deal with the determination of the optimal inventory management parameters.

9.4 Inventory Control Policies

Kennedy *et al.* (2002) present an interesting review of recent literature on spare parts inventories. The proposed models are, mainly, traditional inventory management models and their extensions. According to Silver *et al.* (1998), the main objective of an inventory analysis is to answer the three following questions:

1. How often should the inventory status be determined (control policy)?
2. When should the item be ordered for restocking (order instant)? and
3. How much of the item should be requested at order instant (order quantity)?

In order to find an answer to these questions, it is required that the decision maker sets the following conditions:

1. What is the importance (or criticality) of each item in consideration?
2. Does the inventory position have to be checked continuously or periodically?
3. Of which type the inventory policy should be? and
4. What are the service level targets and costs?

Since all items do not have the same importance (or criticality for spare parts) and due to the fact that a huge number of different parts are kept in stock, while resources are limited, it is usual to adopt decision rules that classify all the items in a limited number of manageable groups. Several classifications methods are available in the literature: Pareto or A-B-C; multicriteria (Braglia *et al.* 2004); variance partition (Eaves and Kingsman, 2004 Williams, 1984), *etc.* Independent of the method used, it is recommended to limit the number of classes to three up to five.

In general, an A-B-C type classification is used: A items receive most personalized attention, strict control policies and have priority over B and C items.

The decision maker has to choose between the continuous and the periodic review strategies. With the continuous review strategy, the stock level is “almost” always known. With the periodic review, the stock level is determined at predetermined instants kR ($k=1, 2, 3, \dots$). The major advantage of continuous review is that it requires less safety stock (hence, lower holding costs) than the periodic review, to provide the same service level (Silver *et al.* 1998).

The main inventory management parameters are the order point (s), the order-up-to level (S), the review period (R) and the economical order quantity (Q). Table

9.4, adapted from Silver *et al.* (1998), proposes an inventory control system selection guide.

Table 9.4. Inventory control system selection guide

Classification	Continuous review	Periodic review
A items	$(S-1, S)$ or (s, Q)	(R, s, S)
B items	(s, Q)	(R, S)
C items	(s, S) or EOQ	(R, S) or EOQ

Several inventory control models taking into account the spare parts features such as long lead-time, slow and random demand, risks of shortage, and obsolescence are examined below. For each considered case, the expressions of the total cost, the order point and the quantity to order are given.

9.4.1 Model with Known and Constant Demand and Lead-time (EOQ Model)

This particular case is the well known Wilson model. The economic order quantity Q^* and the cycle duration T^* are given by

$$Q^* = \sqrt{\frac{2AD}{h}}; \quad T^* = \sqrt{\frac{2A}{hD}}$$

where A is the order cost, h is the holding cost per unit per period and D is the annual demand.

Hadley and Whitin (1963) and Silver and Peterson (1985) have shown that this result was cost insensitive to errors in parameters estimation. This explains why the model is commonly implemented in many commercial software packages despite its shortcomings (see Lee and Nahmias, 1993).

For most practical applications, the value of Q^* is high enough to allow for rounding to the nearest integer without impacting the total cost. In the case of expensive spare parts, and especially the insurance-type spare parts, it is interesting to consider the discrete nature of the demand. The expression of the total cost $TC(Q)$ is then given by Hadley and Whitin (1963):

$$TC(Q) = CD + A \frac{D}{Q} + \frac{h}{2}(Q-1)$$

where C is the acquisition cost of each item or spare part. Q^* is the smallest integer which minimizes $TC(Q)$.

9.4.2 Model with Constant Demand and Perishable Items

Several extensions of the Wilson formula have been proposed. An interesting one is the model dealing with perishable goods for it applies very well to the kind of spare parts subjected to degradation while held in stock. If the stock decays at a constant rate ε , then the instantaneous stock level $I(t)$ is given by Ghare and Schrader (1963):

$$I(t) = \left(I_0 + \frac{D}{\varepsilon} \right) e^{-\varepsilon t} - \frac{D}{\varepsilon}; \quad I_0 = I(0)$$

The expression of the total cost for acquiring and stocking $TC(Q)$ is given by

$$TC(T) = \frac{A}{T} + CD + \left(C \frac{D\varepsilon}{2} + hD \right) T + hD\varepsilon \frac{T^2}{2}$$

T^* satisfies necessarily the following equation:

$$\frac{dTC(T)}{dT} = 0, \quad \text{for } T = T^*$$

which leads to

$$-A + \left(\frac{CD\varepsilon}{2} + hD \right) T^{*2} + hD\varepsilon T^{*3} = 0$$

Once T^* is obtained, the order quantity Q^* is calculated from the following relation:

$$Q^* = D \left(T^* + \varepsilon \frac{T^{*2}}{2} \right)$$

9.4.3 Model with Random Demand and Lead-time

A priori, the state of the stock is hard to find because the demand and lead-time are random. Many inventory control systems have been proposed. See Hadley and Whitin (1963), Hax and Candea (1984), and Kennedy *et al.* (2002) for more details. The description of the most common ones follows:

- Continuous review systems:
 - (s, Q) policy: when the stock level reaches s , Q units are ordered;
 - (s, S) policy: when the stock level becomes equal or less than s , order up to S ;
 - $(S-1, S)$ policy: each time an item is taken from the stock an order is placed to bring the inventory position back to S .
- Periodic review systems:
 - (s, R) policy: at each review time kR ($k = 1, 2, 3, \dots$), a sufficient quantity is ordered to bring the stock level to S ;
 - (s, S, R) policy: if, at review time kR , the stock level is less than or equal to s a sufficient quantity is ordered to bring the stock level up to S ; otherwise, no order is placed.

The $(S-1, S)$ policy, will be presented below. For the other policies, the reader is referred to Hadley and Whitin (1963), Hax and Candea (1984), and Silver *et al.* (1998).

This $(S-1, S)$ inventory policy also called base-stock, is very useful in the inventory control of A items and particularly for expensive spare parts with lifetime longer than replenishment lead-time. The $(S-1, S)$ inventory policy is a special case of the (s, S) policy. It operates as follows: S spare parts are kept in stock and random independent demands, due to replacement at failure, arrive at a rate of λ per unit time. After each spare part request, one replacement unit is ordered. The replenishment lead-time has a general probability distribution with

mean τ . If the nominal stock S is exhausted before the replacements are received, a penalty cost L is incurred for each demand that must be filled by an emergency order or lost due to shortage. A holding cost h per unit time is incurred for each item in stock. This $(S-1, S)$ inventory system, as described above, is equivalent to an M/G/S/S queue whose steady-state probabilities are known as the truncated Poisson distribution (Smith, 1977):

$$Q_s(j) = \text{Probability}\{j \text{ units in stock} \mid \text{desired stock} = S\}$$

$$Q_s(j) = \frac{(\lambda\tau)^{S-j}}{(S-j)!} \bigg/ \sum_{j=0}^S \frac{(\lambda\tau)^j}{j!}; j = 0, 1, 2, \dots, S.$$

The expected total cost per unit time $TC(S)$ in steady-state, given a desired stock level S , is the sum of the average holding cost and the average penalty cost. The expression of $TC(S)$ is therefore given by

$$TC(S) = h \cdot [S - (1 - p(S)) \cdot \lambda\tau] + \lambda L p(S)$$

where

$$p(S) = Q_s(0) = (\lambda\tau)^S / S!$$

S^* is obtained by solving

$$\frac{dTC(S)}{dS} = 0 \quad \text{for } S = S^*.$$

An approximation of S^* is given by

$$S^* = \lambda\tau + \alpha\sqrt{\lambda\tau}$$

where

$$\alpha = \left[2 \ln \left(1 + \frac{L}{h\tau} \right) \right]^{1/2}$$

Several other models are proposed to determine the optimal inventory management parameters for different variants of the $(S-1, S)$ inventory policy by Dhakar *et al.* (1994), Feeney and Sherbrooke (1966), Karush (1957), Moïnzadeh and Schmidt (1991), Schultz (1987), and Walker (1997).

Maintenance policies have an effect on spare parts demand. Frequent preventive replacements reduce random failures but can generate a waste of resources. It seems then advantageous to coordinate maintenance activities and inventory control policies. Decision models should lead to joint determination of maintenance and provisioning periods. Some of these models are presented in the following section.

9.5 Joint Maintenance and Provisioning Strategies

Most analytical models dealing with maintenance strategies assume that whenever a component is to be replaced, either preventively or after a failure, the required resources and spare parts are available on-hand. This implies, as discussed by Brezavšček and Hudoklin (2003), that these components are highly standardized so that the manufacturer can readily procure them, or that they are so inexpensive that

the owner can store large spare parts inventories to protect from failures. In real life these assumptions are rarely satisfied, since most failure prone components are expensive, highly customized and with non-negligible procurement. It is then interesting to combine maintenance and provisioning policies to find an efficient joint policy to perform maintenance actions with the spare parts that have been consequently provisioned. Moreover, recent studies have proven the superiority of joint optimization of maintenance and inventory policies over sequential (or separate) optimization of maintenance and inventory policies (see Acharya *et al.* 1986; Armstrong and Atkins, 1996; Brezavšcek and Hudoklin, 2003; Chelbi and Aït-Kadi, 2001). This section is devoted to the study of these joint strategies which are separated in two different groups: those dealing with the procurement of only one spare part per order (one-unit provisioning) and one dealing with multiple spare parts per order (batch provisioning).

9.5.1 Joint Replacement and Ordering Policy for a Spare Unit (One Unit Provisioning)

Joint replacement and ordering policies for a spare unit are recommended for type A items. Two models will be presented: a basic model without preventive replacement, and a model with preventive replacement.

9.5.1.1 The Basic Model Without Preventive Replacement

Suppose a time t has elapsed since the original unit was put in use. Its replacement spare part is to be ordered at time instant W so that it will be delivered at time $W+L$, where L is assumed constant (see Figure 9.4). A holding cost h is incurred to store the part. A shortage cost π is charged for each shortage period.

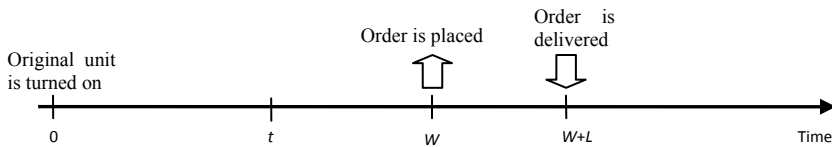


Figure 9.4. Ordering and replacement cycle

The expected total cost $TC(W)$ is the sum of the average holding cost and the average shortage cost (Mitchell 1962):

$$TC(W) = \pi \int_t^{W+L} \frac{f(x)}{R(t)} dx + h \int_{W+L}^{\infty} \frac{R(x)}{R(t)} dx$$

The instant W^* which minimizes $TC(W)$ satisfies

$$\frac{dTC(W)}{dW} = 0 \text{ for } W = W^*$$

which is equivalent to finding W such as

$$r(W+L) = \frac{h}{\pi}$$

where $r(t) = f(t)/R(t)$ is the failure rate.

9.5.1.2 Optimal one-unit Ordering Policy for Preventively Replaced Systems

In the previous model, the ordered part is used only when the operating item breaks down. Preventive replacement was not considered. However, for systems with increasing failure rate, preventive actions are necessary to reduce the replacements costs or avoid catastrophic breakdowns. The following model, proposed by Dohi *et al.* (1996), discusses a generalized ordering policy with time-dependent delay structure, in which when an ordered spare is delivered after a lead time, it is put into the inventory if an original unit is still operating, and the original one is replaced/exchanged by the spare in stock when the original one fails/passes a pre-specified time, whichever occurs first.

The original unit begins operating at time 0, and the planning horizon is infinite. If the original unit does not fail up to a pre-specified time t_0 , the regular order for a spare is made at the time t_0 and after a deterministic lead-time L the spare is delivered. Then, if the original unit has already failed, the delivered spare takes over its operation immediately. But if the original unit is still operating, the spare part is put into the inventory, and the original one is replaced or exchanged by the spare in the inventory when the original one fails or passes a pre-specified time interval $t_1 - L$ ($t_1 \in [L, \infty]$) after the spare is delivered, whichever occurs first. On the other hand, if the original unit fails before the time t_0 , an expedited order is placed immediately at the failure time t and the spare takes over its operation as soon as it is delivered after a lead time $L_e(t)$. Each unit has lifetime density function $f(t)$.

The costs considered are the following: a cost π per unit time is incurred for the shortage period of the original unit; h is the holding cost per unit inventory period of a spare; and costs A_1 and A_2 are incurred for all expedited and regular orders, respectively.

By deriving the expected total cost per unit time in the steady-state $K(t_0, t_1)$, Dohi *et al.* obtained the following theorem governing the optimal ordering policies: for any ordering time t_0 , the optimal allowed inventory time for a spare t_1^* which minimizes the expected total cost $K(t_0, t_1)$

$$t_1^* \rightarrow \infty \text{ if } N(t_0) \leq 0;$$

$$t_1^* \rightarrow L \text{ if } N(t_0) \geq 0.$$

where

$$\begin{aligned} N(t_0) = & \pi \left[\int_0^{t_0} l_e(t) \cdot F(t) dt - \int_{t_0}^{t_0+L} F(t) dt + (L - L_e(t_0)) F(t_0) \right] \\ & + h \cdot \left[\int_0^{t_0} (1 - l_e(t)) R(t) dt + (L_e(t_0) - L) F(t_0) \right] + \\ & [L - L_e(t_0) + L_e(0)] - A_1 \cdot F(t_0) - A_2 \cdot R(t_0). \end{aligned}$$

and

$$l_e(t) = dL_e(t)/dt.$$

This important result means that we should only consider either the extreme case where the delivered spare is put into the inventory until the original unit fails ($t_1^* \rightarrow \infty$), or the other case where the spare takes over the operation as soon as it is

delivered ($t_1^* \rightarrow L$). For each case, the optimal ordering instants t_0^* are given in Dohi *et al.* (1996).

Several other variants of this 1-unit ordering policy have been proposed by Kaio (1988), Osaki *et al.* (1981), Park and Park (1986), Sheu *et al.* (1992), and Thomas and Osaki (1978). A comprehensive bibliography on this topic can be found in Dohi *et al.* (2006).

9.5.2 Joint Replacement and Multiple Spare Parts Ordering Policy (Batch Provisioning)

For some systems, the replacement process is such that several spare parts are required during a single replenishment cycle. Therefore, a batch of multiple spare parts is ordered at once. The resulting joint maintenance and provisioning strategy usually combines age or block replacement policy with well-known inventory policies such as (s, Q) , (R, S) , and (s, S) . Acharya *et al.* (1986) have studied a joint block replacement under (R, S) inventory control policy. A mathematical model is proposed to derive the optimal block replacement strategy T and the inventory control parameters R and S which minimize the expected total cost. They assume a negligible lead-time and consider that spare parts are always available at preventive maintenance instants. The optimal parameters are determined for a given lifetime distribution. As an extension of Acharya's model, Chelbi and Aït-Kadi (2001) proposed a procedure to determine the optimal strategy for a general lifetime distribution. Lately, Brezavšček and Hudoklin (2003) presented a similar model for a system with k identical units operating simultaneously with non-negligible procurement lead-time. The optimal strategy is derived from a total cost function.

Few other joint replacement and provisioning models are devoted to the availability maximization. Al-Bahi (1993) considers a system with k identical units operating with constant failure rate λ_f and supported by a (s, Q) spare parts provisioning policy. The optimal parameters are derived to ensure minimum inventory costs without degrading the spare availability over the inventory cycle. Sarker and Haque (2000) studied a joint block replacement and (s, S) provisioning policy using a simulation model. Recently, Diallo *et al.* (2008) have proposed a mathematical model for the maximization of the system's availability under joint preventive maintenance and (s, Q) spare parts provisioning strategy.

The model proposed by Chelbi and Aït-Kadi (2001) combines block replacement policy with a (R, s) inventory review, where s is the order point and R is the review period ($R = kT$). The block replacement policy suggests using new components to perform replacements at failure and at pre-determined instants T , $2T$, $3T$, ... regardless of the state and age of the system. The system is made up of n independent and identically distributed components. The expected total cost per unit time over an infinite horizon $B(s, T, R)$ is the sum of the replacement, holding, order and shortage costs. C_C and C_P are the costs for performing a replacement at failure and a preventive replacement respectively. The probability density function of the demand during the lead-time $g(x)$ is assumed to be normally distributed. The expression of $B(s, T, R)$ is given by

$$B(s, T, R) = n \left[\frac{C_C M(T) + C_P}{T} \right] + n \frac{h}{T} \int_0^T \left[\frac{k+1}{2} M(T) - M(t) \right] dt + \frac{A}{kT} \\ + h(s - knM(T)) - (\pi - h)s \int_s^\infty g(x)dx + (\pi + h) \int_s^\infty xg(x)dx$$

For a given k , the optimal strategy (s^*, T^*, R^*) , if it exists, satisfies necessarily the following system of equations:

$$\frac{\partial B(s, T, R)}{\partial T} = 0 \quad \text{for } T = T^*, R = R^*, s = s^*$$

$$\frac{\partial B(s, T, R)}{\partial s} = 0 \quad \text{for } T = T^*, R = R^*, s = s^*$$

The results displayed in Table 9.5 are obtained for the following set of data:

$C_C = \$70$, $C_P = \$20$, $h = \$1$, $A = \$50$, $\pi = \$20$, $n = 150$ units.

Table 9.5. Optimal strategy for items with Weibull (2,4) distributed lifetime

k	T^*	$R^*=kT^*$	s^*	$B(s^*, T^*, R^*)$
1	0.10	0.10	31	31475.2
2	0.10	0.20	57	31228.3
3	0.10	0.30	82	31171.5
4	0.10	0.40	107	31128.8
5	0.10	0.50	131	31128.4
$k^*=6$	0.10	0.60	156	31113.4
7	0.10	0.70	180	31125.7
8	0.10	0.80	204	31118.3
15	0.10	1.50	371	31188.2
25	0.10	2.50	608	31297.3

The models presented in this section account for maintenance actions in the determination of the inventory control parameters. This joint-optimization of both maintenance and inventory policies yields substantial savings. However, these models consider that the replacements are carried out with new components. How would the use of reconditioned (or used) items, having a given age and lower cost, impact the maintenance strategies? How should those reconditioned parts be selected and used? These reconditioned components are acquired from external suppliers or recovered during preventive maintenance actions or after the repair of the components removed following a breakdown. Section 9.6 will deal with reconditioned components and their impact on maintenance and inventory control policies.

9.6 Inventory and Maintenance Policies for Reconditioned Spare Parts

According to Fleischmann *et al.* (2003), recovered spare parts can cost up to 80% less than new ones. However, having already been in use (age $x > 0$), those parts are usually less reliable than new ones and come from less regular provisioning sources. The low acquisition cost gain can easily be overturned by the high replacement costs yielded by more frequent failures due to using less reliable aged parts. It is then required to determine the adequate age of the recovered parts to be used, considering their remaining lifetime, their reliability and the costs related to the additional replacement actions.

9.6.1 Age of Recovered Parts to be Used for Replacement Actions

A component is said to have age x ($x > 0$) if it has been operating without failure during x units of time. If $x = 0$, then the component is new. If $f(\cdot)$ denotes the lifetime distribution of a new component then, the lifetime distribution function $f_x(\cdot)$ for a component of age x is given by

$$f_x(t) = \frac{f(x+t)}{R(x)}.$$

The reliability function of this component of age x is

$$R_x(t) = \frac{R(x+t)}{R(x)}.$$

For a component with non-decreasing failure rate: $R_x(t) \leq R(t)$, $\forall x \geq 0; \forall t \geq 0$. The distribution function of such a component is said to be NBU (new better than used) (see Barlow and Proschan, 1981; Bryson and Siddiqui, 1969). It can also be shown that its mean residual lifetime decreases with age.

Another useful metric is the average number of renewals $M_u(T)$ in $[0, T]$ when replacements are carried out with used components.

If the original part is new and the replacements are performed using components with age x , then

$$M_u(T) = \int_0^T [1 + M_x(T-y)] f(y) dy$$

Where $M_x(T)$ satisfies the renewal equation

$$M_x(T) = \int_0^T [1 + M_x(T-y)] f_x(y) dy$$

$$M_x(T) = \sum_{n=1}^{\infty} F_x^{(n)}(T)$$

where $F_x^{(n)}$ is the n -fold convolution of F_x with itself.

If the original part and the components used for replacements have all age x , then

$$M_u(T) = M_x(T)$$

$$M_u(T) = \sum_{n=1}^{\infty} F_x^{(n)}(T)$$

If for a mission of duration T , the reliability target is R_L then the age x of the reconditioned part to use is such as $R_x(T) \geq R_L$, which is equivalent to finding x such that

$$\frac{R(x+T)}{R(x)} \geq R_L$$

If the determination of the age is based on the least mean residual lifetime (MRL) to achieve, the reconditioned part will have age x such that

$$\frac{\int_x^{\infty} R(t) dt}{R(x)} \geq MRL.$$

The following model deals with the determination of the optimal age of the components used to carryout replacements at failure considering the acquisition and replacements costs. The goal is to find a trade-off between the low acquisition costs of reconditioned parts and their higher total replacement cost.

We consider a system exploited over a horizon of length T ($T > 0$). Denote by $TC_N(T)$ the total cost for acquiring and performing replacements at failure with new parts. $TC_U(T)$ denotes the total cost for acquiring and performing replacements at failure with parts with age x .

Each new part costs C . Each reconditioned part costs $(C - C_{\min})e^{-bx} + C_{\min}$ where b is the decrease rate of the component acquisition cost with respect to its age. $C_{\min}$ is the lowest cost at which a reconditioned part can be bought. Replacements at failure are performed at a cost C_R . The goal is to find the optimal age x , of the reconditioned parts to be used, provided that $TC_U(T)$ is less or equal to $TC_N(T)$. In both cases, the original component is new (see Figure 9.5).

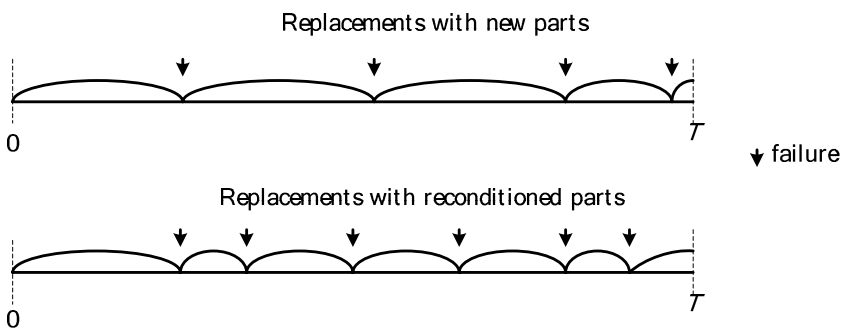


Figure 9.5. Replacements occurrences

We have

$$TC_N(T) = (C + C_R) \cdot M(T)$$

$$TC_U(T) = \left[(C + C_{\min}) \cdot e^{-bx} + C_{\min} + C_R \right] \cdot M_u(T)$$

Therefore, we should find x such that

$$(C + C_{\min}) \cdot e^{-bx} + C_{\min} + C_R - (C + C_R) \cdot \frac{M(T)}{M_u(T)} \leq 0$$

for $T > 0$ ($M_u(T) \neq 0$)

Denoting the left-hand side of the inequality by $\varphi(x)$, we get

$$\varphi(x) \leq 0 \quad \text{for} \quad T > 0$$

where

$$\varphi(x) = (C + C_{\min}) \cdot e^{-bx} + C_{\min} + C_R - (C + C_R) \cdot \frac{M(T)}{\int_0^T [1 + M_x(T-y)] \cdot f(y) dy}$$

Due to the complexity of the expression of $M(T)$, solving the inequality is cumbersome for general lifetime distributions. Consider the following lifetime distribution:

$$f(t) = te^{-t};$$

thus,

$$M(T) = 0.5T + 0.25e^{-2T} - 0.25$$

$$M_x(T) = \frac{-x^2 - 2x - 1 + (x^2 + 3x + 2)T + (x^2 + 2x + 1)e^{\frac{(x+2)T}{x+1}}}{x^2 + 4x + 4}$$

and

$$\begin{aligned} \varphi(x) = & (C + C_{\min}) \cdot e^{-bx} + C_{\min} + C_R \\ & - (C + C_R) \cdot \left(\frac{(0.5T + 0.25e^{-2T} - 0.25) \cdot (x^2 + 4x + 4)}{-x^2 - 2x - 1 + (x^2 + 3x + 2)T + (x^2 + 2x + 1)e^{\frac{(x+2)T}{x+1}}} \right) \end{aligned}$$

Solving

$$\varphi(x) \leq 0$$

with

$$C = \$500, C_R = \$200, C_{\min} = \$200, b = 0.7, \text{ and } T = 10$$

yields

$$x \in [1.69, 9.64].$$

which means that it is economical to perform replacements at failure with parts aged between 1.69 and 9.64.

A representation of the function $\varphi(x)$ is shown in Figure 9.6. The maximum gain is achieved whenever parts with age $x^* = 3.89$ are used.

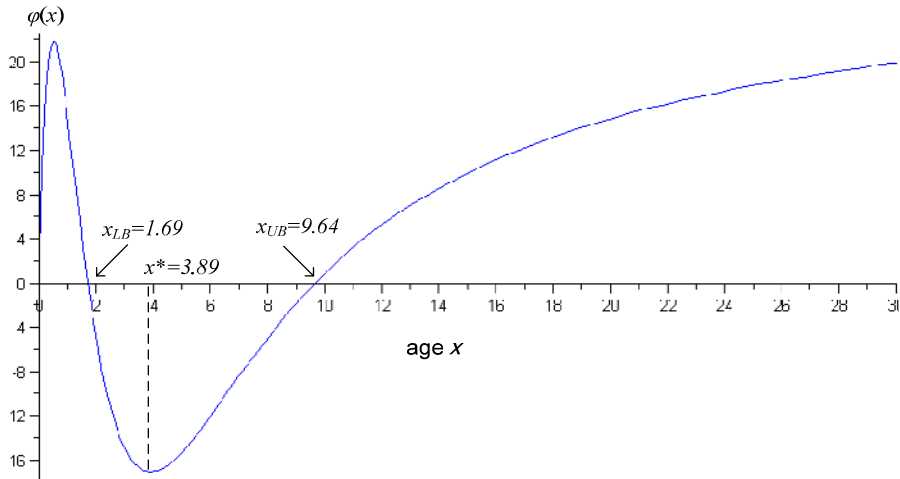


Figure 9.6. Function $\varphi(x)$

If there is an interval where $\varphi(x)$ is convex and negative, then there is a unique optimal age x^* which maximizes the savings procured by performing replacements with used or reconditioned items. This optimal age x^* is solution of

$$\frac{d\varphi(x)}{dx} = 0 \quad \text{for} \quad x = x^*$$

This illustrative example shows that used spare parts can efficiently be integrated in a maintenance strategy, provided that reliable suppliers of those reconditioned parts can be found. In the coming years, it will be less difficult to access the market of reconditioned items, because of the recovery and reuse legislations, Extended Producer Responsibility acts, and environmental requirements for sustainable development that are being implemented in many countries around the world. The automobile and electronics industries have already undertaken pioneering actions in recovering and reusing their end-of-life products as reported by Guide and Van Wassenhove (2001), Guide *et al.* (2005), and Hormozi (1997).

If no external suppliers can be found, components recovered during preventive replacement actions or after failure can be repaired or reconditioned and be used. Several models have been proposed to deal with this internal source of reconditioned parts. Bhat (1969) proposed a policy where new items are used for preventive replacements at instants kT ($k=1, 2, 3, \dots$), and used parts with age T are used for replacements at failure. Tango (1978, 1979) divides the replacement cycle T in two intervals: $[(k-1)T, kT-\delta]$ and $[kT-\delta, kT]$. Replacements are carried-out with new parts at kT and for failures occurring in $[(k-1)T, kT-\delta]$. For failures occurring in $[kT-\delta, kT]$, replacements are done with used parts of age T . Murthy and Nguyen (1982) proposed a policy where all the components recovered during preventive replacements are used, regardless of their ages, to carryout replacements at failure. Aït-Kadi *et al.* (2003b) proposed a policy where preventive replacements

at instants kT ($k=1, 2, 3, \dots$) are carried-out with new items and replacements at failure are performed with used parts of age x , where x is a decision variable to be determined. Other models integrating minimal repair and used parts can be found in Aït-Kadi and Cl  roux (1988), Aït-Kadi *et al.* (1990), and Nakagawa (1981, 1982).

9.6.2 Review of Inventory Control Policies with Random Returns

Be it a manufacturing company which acquires used parts from external suppliers or a company engaging in product recovery for resale purposes, both must face limited and random availability of used components. They must then fulfil their needs from traditional new-parts suppliers (see Teunter 2001, 2004). New inventory control parameters must then be derived to account for the existence of two supplying sources. Fleischmann (2001), Nahmias and Rivera (1979), and De Brito and Dekker (2003) present extensive reviews of inventory control policies with return loop also called closed-loop models.

9.7 Collaborative Management of Spare Parts

Despite the complexity of the spare parts inventory problem, many inventory system management improvements can be achieved. Significant gains can be yielded through the use of the internet and the new connective technologies such as RFID (radio-frequency identification). Lead-time reduction, rigorous ordering and stock monitoring, wider access to suppliers from all over the world, better prices, eased communications with suppliers, improved access to updates and user guides are among the advantages provided by the internet and the new communication technologies.

9.7.1 Access to Documentation and Knowledge Bases

Many manufacturers set up websites where their customers can access the technical documentation and information on the equipment they bought. Equipment updates, safety modifications, service packs, re-design updates or even recall information are made available through those websites. The customers can download up to date information or software as soon as they are released, instead of having to wait for them to be delivered by mail. Customers can subscribe to technical newsletters dedicated to their equipment or access discussion forums where they can report equipment problems and get answers both from the manufacturers and other users. Cope (2000) reports that Pratt & Whitney Online Services allows its customers to access their online parts catalogues, training and user guides, diagnosis tools, and fleet performance data. Boeing offers, on its website myboeingfleet.com, access to more than 6 million spare parts.

9.7.2 Lead-time Reduction

One aspect of the inventory control that has been deeply affected by the internet is the procurement lead-time which has been significantly shortened by the online purchase transactions as described by Cross (2000) and Westerkamp (1998). By reducing the procurement lead time, the safety stocks are scaled down. In the traditional ordering process, the storekeeper must fill an order form which is sent (by mail or fax) to the manufacturer or supplier. Upon reception of the order form, the supplier manufactures, packs and ships the requested quantities. With the online catalogues, the spare parts ordering procedures are simplified. A series of mouse clicks on an image or a drop-down menu is sufficient to select the desired component, hence avoiding reference number transcription errors. After confirmation of the purchase, a cascade of logistic operations is triggered and ends with the delivery of the component after few hours or days. The order transmission phase to the supplier is almost instantaneous. With the increasing number of online transactions, the parcel delivery systems have improved and their costs have decreased. Several logistic service providers make it possible to track ordered items and thus better plan their reception and the ensuing operations.

9.7.3 Virtually Centralized Spare Parts Stock (Inventory Pooling)

The Internet and data exchange technologies allow several forms of collaboration between companies. This e-collaboration between companies is already in application in the supply chain of many manufactured products (see Holmström, 1998; Huiskonen, 2001; Kilpi and Vepsäläinen 2004). This e-collaboration can be horizontal when companies, at the same echelon, work together. This is the case with inventory pooling, the joint replenishment (also known as order pooling), vehical sharing, *etc.* Vertical e-collaboration is applied, when organizations from different echelons become partners. This is the case with the Vendor-managed inventory (VMI) when the supplier or manufacturer manages its customers' stocks.

Inventory pooling of spare parts can be real (physical) with several companies served from one centralized store (see Figure 9.7a) or virtual (Schneider and Watson, 1997) when each company keeps its share of the joint-stock on its premises but can dispatch or receive parts from the other partners (see Figure 9.7b) (Dong and Rudi, 2004; Kukreja and Schmidt, 2005). Excellent sharing of the stock levels information is required for the system to operate.

It has been proved that inventory pooling always reduces the total inventory cost (statistical economies of scale). This concept was modelled by Eppen (1979). He proved that the total inventory cost of a decentralized system TC_d exceeds the total cost in a centralized system TC_c .

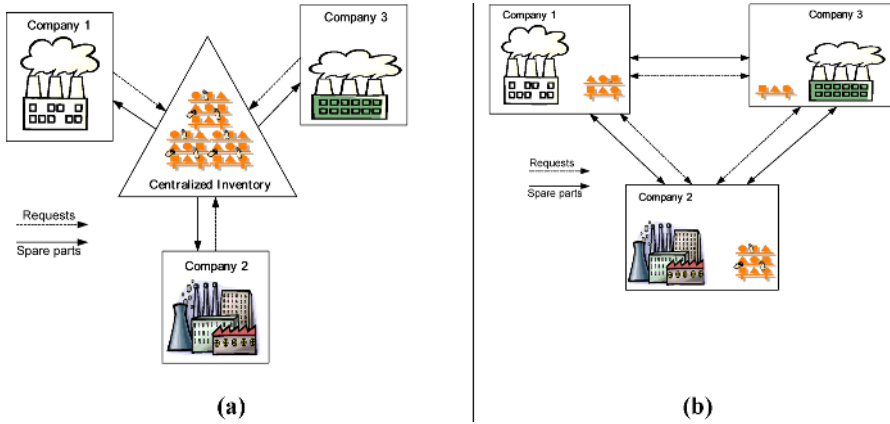


Figure 9.7. Centralized stock: (a) physically; (b) virtually

Considering a group of N partners, if the spare parts demand d_i for each partner P_i ($i=1, 2, \dots, N$) has a normal distribution with mean $E[d_i] = \mu_i$ and variance σ_i^2 , Eppen (1979) has shown that

$$TC_c = K \sqrt{\sum_{i=1}^N \sigma_i^2} \quad \text{and} \quad TC_d = K \sum_{i=1}^N \sigma_i$$

Hence

$$TC_c \leq TC_d$$

where K is a function of holding costs, shortage costs and demand distribution.

Centralization is less expensive because higher than average demand for a partner is absorbed or compensated by less than average demand experienced by other partners. Virtually pooled inventory is the core principle of the commercial business SparesFinder (www.sparesFinder.net).

Figure 9.8 depicts the variation of the number of spare parts, required to achieve a given service level, as a function of the total number of machines owned by the partners. Figure 9.8 is obtained from Equation 9.1 established in Section 9.2.2.

It should be noted that if inventory pooling always reduces the total inventory costs as shown by Cherikh (2000) and Eppen and Schrage (1981), it does not, in general, guarantee a reduction in stock level: this is called the “inventory pooling anomaly” (see Chen and Lin, 1990; Dong and Rudi, 2004; Gerchak and Mossman, 1992; Yang and Schrage, 2003; Zhang, 2005). Recently, Yang and Schrage (2009) investigated the conditions that cause pooling to increase inventory levels. They showed that this anomaly can occur with any right skewed demand distribution.

After analyzing thousands of industrial cases, Hillier (2002) came to the conclusion that benefits yielded by risk pooling are less than the benefits of joint replenishment.

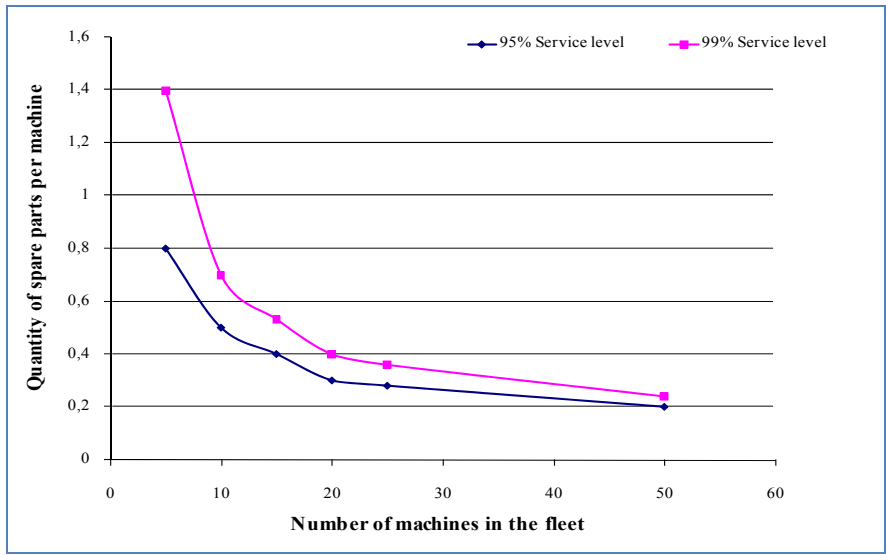


Figure 9.8. Variation of the number of spare parts according to the total number of machines and service level

9.7.4 Joint Replenishment of Spare Parts

Joint ordering spare parts by several partners can yield substantial economies of scale on the purchase and transportation costs as shown by Nielsen and Larsen (2005). Let us consider N companies deciding to join their spare parts ordering. Because the shortage, storage, and transportation costs are not identical for all the partners, we consider that there are N SKU (stock-keeping units) even if it is the same spare part that is ordered. The problem is then to coordinate the order of the same spare part for N companies. By considering N different SKU, this problem becomes equivalent to the traditional joint replenishment problem (JRP) where the problem is to coordinate the order of N different items for one location or company. One partner, denoted company 1, is designated to centralize all the orders, place the joint-order, take delivery of the requested spare parts, and distribute individual batches to the other companies (see Figure 9.9). The order cost is composed of a major cost A and N minor costs a_i ($i=1, 2, \dots, N$). The major cost includes all the costs incurred by company 1 in the process of ordering the spare parts, receiving, inspecting and separating them into individual batches. The minor cost a_i is the cost incurred by partner i whenever he decides to place an order with the others. This minor cost includes all the costs incurred by the partner for transmitting his order quantity to company 1, and the transportation cost to bring the spare parts from company 1. The JRP has been largely covered in the literature. A review of deterministic and stochastic models is presented by Goyal and Satir (1989).

To conclude this section, we propose a set of actions that can be undertaken to reduce the total inventory and maintenance cost. This total inventory and maintenance cost is the sum of the ordering, acquisition, holding, shortage, and

replacement costs. Table 9.6. lists the potential cost-reduction initiatives for each cost type.

Table 9.6. Cost-reduction initiatives for each cost type

Cost	Definition	Cost-reduction actions
A	Ordering cost: it includes all the costs for preparing and placing the order, follow-up, and reception of the ordered articles	Rationalize and join orders; Use e-commerce tools for ordering and monitoring orders; Develop partnership with suppliers (Vendor-managed Inventory, production & delivery coordination, ...etc.); Develop collaborations with other users
C	Acquisition cost: cost to buy the item (variable cost)	Regroup orders to benefit from scale savings; Monitor and search for temporary reduction offers (subscription to OEMs technical newsletters); Regularly research new suppliers in order to extend suppliers list and lower purchase costs; Recourse to reconditioned spare parts
h	Holding cost: variable cost including all the expenses incurred by the presence of an item in stock (rent, insurances, taxes, interests, wages, etc.). The holding cost of a unit over a given horizon accounts for 20 – 60% of its acquisition cost	Reduce the quantities kept in stock through risk-pooling techniques; Reduce stocking period (just-in-time, determination of the optimal ordering instant); Recourse to reconditioned spare parts
π	Shortage cost: sum of all costs incurred following a shortage. It can be difficult to evaluate but generally includes the cost related to the loss of capacity, customer compensation, lost customers, backlogging or emergency delivery of the quantities in shortage	Keep an emergency supplying source (lateral transshipment, Overnight delivery, etc.); Reduce lead-times (online ordering, rapid machining); Parts interchange-ability and commonality
MC	Maintenance cost: it is the sum of the costs for all the actions carriedout in order to maintain or to restore the equipment in a good operating condition	Improve employees training; Prepare and organize maintenance actions (computerized maintenance management systems CMMS); Link the CMMS system with the procurement system; Access to online or electronic documentation (OEM web site, internet-based customer care or assistance)

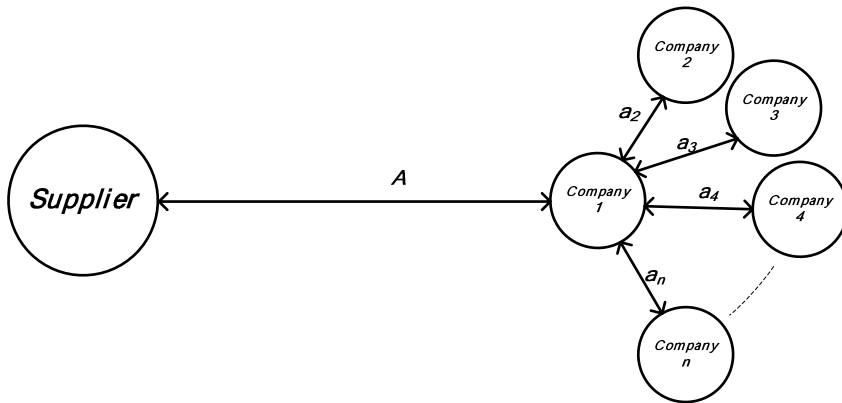


Figure 9.9. Order cost repartition

9.8 Conclusion

An integrated approach for the identification and the management of spare parts has been proposed. We described a methodology for the identification of the components for which spare parts should be kept. For each spare part, analytical models are presented for the determination of the quantities required over a given horizon. Inventory management models have then been derived for the provisioning of spare parts needed to carryout the preventive and corrective maintenance actions. Several factors affecting the system performance, such as provisioning leadtime, random demand, and perishability, are considered in the selected mathematical models aiming to determine the inventory management parameters. The contribution of reconditioned spare parts is also investigated. Mathematical models for the determination of the optimal age of the reconditioned spare parts are derived. We also investigated how the integration of adequate information technology may contribute to the improvement of the spare parts stock management system.

Because the machines and their operating environment tend to change over time, it is judicious to frequently update management parameters and decision variables to account for technical, economic and strategic changes. It is also worth implementing maintenance procedures for the replacement parts, when they are in storage through appropriate inspections and control of the environment conditions (humidity level, greasing, repositioning by rotation or flipping, *etc.*).

References

- Acharya D, Nagabhushanam G, Alam SS (1986) Jointly optimal block-replacement and spare provisioning policy. *IEEE Trans Reliab* R-35: 447–451
- Aït-Kadi D, Chelbi A (1998) Procédure numérique de calcul de la fonction de renouvellement. Tech Rep, Université Laval, Québec.

- Aït-Kadi D, Clérout R (1988) Optimal block policies with multiple choice at failure. *Nav Res Logist* 35: 99–110.
- Aït-Kadi D, Beaucaire C, Clérout R (1990) A periodic maintenance model with used equipment and random minimal repair. *Nav Res Logist* 37: 855–865.
- Aït-Kadi D, Diallo C, Chelbi A (2003a) Spare parts identification and provisioning models. *J Decis Syst* 12: 47–65
- Aït-Kadi D, Jamali MA, Artiba A (2003b) Optimal periodic replacement strategy using new and used items. *J Decis Syst* 12: 67–77.
- Al-Bahi AM (1993) Spare provisioning policy based on maximization of availability per cost ratio. *Comput Ind Eng* 24: 81–90
- Armstrong MJ, Atkins DR (1996) Joint optimization of maintenance and inventory policies for a simple system. *IIE Trans* 28: 415–424
- Barlow RE, Proschan F (1965) *Mathematical theory of reliability*. John Wiley and Sons, New-York
- Barlow RE, Proschan F (1981) *Statistical theory of reliability and life testing: probability models*. To Begin With, Silver Spring
- Bhat BR (1969) Used items replacement policy. *J Appl Probab* 6: 309–318
- Braglia M, Grassi A, Montanari R (2004) Multi-attribute classification method for spare parts inventory management. *J Qual Maint Eng* 10: 55–65
- Brezavšek A, Hudoklin A (2003) Joint optimization of block-replacement and periodic-review spare-provisioning policy. *IEEE Trans Reliab* 52: 112–117
- Brown RG (1977) *Materials management systems*. Wiley-Interscience, New York
- Bryson MC, Siddiqui MM (1969) Some criteria for aging. *J Am Stat Assoc* 64: 1472–1483
- Burton RW, Jacquette SC (1973) The initial provisioning decision for insurance type items. *Nav Res Logist* 20: 123–146
- Chelbi A, Aït-Kadi D (2001) Spare provisioning strategy for preventively replaced systems subjected to random failures. *Int J Prod Econ* 74: 183–189
- Chelbi A, Aït-Kadi D (2002) Classement des équipements par ordre de priorité pour la maintenance. *J Decis Syst* 11: 911–08
- Chen MS, Lin CT (1990) An example of disbenefits of centralized stocking. *J Oper Res Soc* 41: 259–262
- Cherikh M (2000) On the effect of centralization on expected profits in a multi-location newsboy problem. *J Oper Res Soc* 51: 755–761
- Cope J (2000) Pratt & Whitney to link parts purchase to service docs. *Computerworld* 34:12
- Cross GJ (2000) How e-Business is transforming supply chain management. *J Bus Strateg* 21: 36–39
- De Brito MP, Dekker R (2003) Modelling product returns in inventory control - exploring the validity of general assumptions. *Int J Prod Econ* 81: 225–241
- Dhakar TS, Schmidt CP, Miller DM (1994) Base stock level determination for high cost low demand critical repairable spares. *Comput Oper Res* 21: 411–420
- Diallo C (2006) *Développement d'un modèle d'identification et de gestion des pièces de rechange*. PhD. Diss, Université Laval, Québec
- Diallo C, Aït-Kadi D (2007) A simple approximation to compute the mean number of failures for a repairable system and application to the availability determination. *Proc Congr Int Génie Ind, Trois-Rivières, Canada*
- Diallo C, Aït-Kadi D, Chelbi A (2008) (s,Q) spare parts provisioning strategy for periodically replaced system. *IEEE Trans Reliab* 57: 134–139
- Dohi T, Kaio N, Osaki S (1996) Optimal ordering policies with time-dependent delay structure. *J Qual Maint Eng* 2: 50–62
- Dohi T, Ashioka A, Kaio N, Osaki S (2006) Statistical estimation algorithms for repairs-time limit replacement scheduling under earning rate criteria. *Comput & Math Appl* 51: 345–356

- Dong L, Rudi N (2004) Who benefits from transshipment? Exogenous vs. endogenous wholesale prices. *Manag Sci* 50: 645–657
- Dorsch J (1998) Spare parts create profits in gear market. *Electron News* 44: 34–38
- Eaves AHC, Kingsman BG (2004) Forecasting for the ordering and stock-holding of spare parts. *J Oper Res Soc* 55: 431–437
- Eisenhower SW, Bott TF, Jackson JW (2002) Prioritizing the purchase of spare parts using an approximate reasoning model. *Proc Annual Reliab & Maintainab Symp* 2002: 20–25
- Eppen GD (1979) Effects of centralization on expected costs in a multi-location newsboy problem. *Manag Sci* 25: 498–501
- Eppen GD, Schrage L (1981) Centralized ordering policies in a multi-warehouse system with lead-times and random demands. In Schwarz LB (ed) *Multi-level production/inventory control systems: theory and practice*. North-Holland, Amsterdam
- Feeney GJ, Sherbrooke CC (1966) The (s-1,s) inventory policy under compound Poisson demand. *Manag Sci* 12: 391–411
- Fleischmann M (2001) *Quantitative models for reverse logistics*. Springer-Verlag, Berlin
- Fleischmann M, van Nunen J, Gräve B (2003) Integrating closed-loop supply chains and spare-parts management at IBM. *Interfaces* 33: 44–56
- Gajpal PP, Ganesh LS, Rajnedran C (1994) Criticality analysis of spare parts using the analytic hierarchy process. *Int J Prod Econ* 35: 293–297
- Gerchak Y, Mossman D (1992) On the effect of demand randomness on the inventories and costs. *oper res* 40: 804–807
- Geurts JHJ, Moonen JMC (1992) On the robustness of 'insurance type' spares provisioning strategies. *J Oper Res Soc* 43: 43–51
- Ghare PM, Schrader GF (1963) A model for an exponentially decaying inventory. *J Ind Eng* 14: 238–243
- Goyal SK, Satir AT (1989) Joint replenishment inventory control: deterministic and stochastic models. *Eur J Oper Res* 38: 2–13
- Gross D (1976) Sensitivity of queueing models to the assumption of exponentiality. *Nav Res Logist Q* 22: 271–287
- Gross D, Kahn HD, Marsh JD (1977) Queueing models for spares provisioning. *Nav Res Logist Q* 24: 521–536
- Guide VDR Jr, Van Wassenhove LN (2001) Managing product returns for remanufacturing. *Prod & Oper Manag* 10: 142–155
- Guide VDR Jr, Muyldermans L, Van Wassenhove LN (2005) Hewlett-Packard company unlocks the value potential from time-sensitive returns. *Interfaces* 35: 281–293
- Hadley G, Whitin TM (1963) *Analysis of Inventory Systems*. Prentice-Hall, Englewood Cliffs
- Haneveld WKK, Teunter RH (1997) Optimal provisioning strategies for slow moving spare parts with small lead times. *J Oper Res Soc* 48: 184–194
- Hax A, Candea D (1984) *Production and inventory management*. Prentice-Hall, Englewood Cliffs
- Hillier MS (2002) Using commonality as backup safety stock. *Eur J Oper Res* 136: 353–365
- Holmström J (1998) Implementing vendor managed inventory the efficient way: a case study of partnership in supply chain. *Prod & Invent Manag J* 1998: 1–5
- Hormozi AM (1997) Parts remanufacturing in the automotive industry. *Prod & Invent Manag J* 38: 26–31
- Huiskonen J (2001) Maintenance spare parts logistics: special characteristics and strategic choices. *Int J Prod Econ* 71: 125–133
- Kaio N (1988) Optimum planned policies with minimal repair. *Microelectron Reliab* 28: 287–293
- Karush W (1957) A queueing model for an inventory. *Oper Res* 5: 693–703

- Kennedy WJ, Patterson JW, Fredendall LD (2002) An overview of recent literature on spare parts inventories. *Int J Orod Econ* 76: 201–215
- Kilpi J, Vepsäläinen APJ (2004) Pooling of spare components between airlines. *J Air Transp Manag* 10: 137–146
- Kukreja A, Schmidt CP (2005) A model for lumpy demand parts in a multi-location inventory system with transshipments. *Comput Oper Res* 32: 2059–2075
- Lee HL, Nahmias S (1993) *Logistics of Production and Inventory*. North Holland, Amsterdam
- Mitchell GH (1962) Problems of controlling slow-moving engineering spares. *Oper Res Q* 13: 23–39
- Moinzadeh K, Schmidt CP (1991) An (S-1,S) inventory System with emergency orders. *Oper Res* 39: 308–321
- Murthy DNP, Nguyen DG (1982) A note on extended block replacement policy with used items. *J Appl Probab* 19: 885–889
- Nahmias N, Rivera H (1979) A deterministic model for a repairable item inventory system with finite repair rate. *Int J Prod Res* 17: 215–221
- Nakagawa T (1981) A summary of periodic replacement with minimal repair at failure. *J Oper Res Soc Jpn* 24: 213–227
- Nakagawa T (1982) A modified block replacement with two variables. *IEEE Trans Reliab* 31: 398–400
- Nielsen C, Larsen C (2005) An analytical study of the Q(s,S) policy applied to the joint replenishment problem. *Eur J Oper Res* 163: 721–732
- Osaki S, Kaio N, Yamada S (1981) A summary of optimal ordering policies. *IEEE Trans Reliab R-30*: 272–277
- Park YT, Park KS (1986) Generalized spare ordering policies with random lead time. *Eur J Oper Res* 23: 320–330
- Sarker R, Haque A (2000) Optimization of maintenance and spare provisioning policy using simulation. *Appl Math Model* 24: 751–760
- Schärlig A (1985) *Décider sur plusieurs critères: Panorama de l'aide à la décision multicritère*. Presses Polytechniques Romandes, Lausanne
- Schneider H, Watson E (1997) An analysis of warehouse and distribution strategies: a case study. *OR Spektrum* 19: 169–175
- Schultz CR (1987) Forecasting and inventory control for sporadic demand under periodic review. *J Oper Res Soc* 38: 453–458
- Sherbrooke CC (1968) Metric: a multi-echelon technique for recoverable item control. *Oper Res* 16: 122–141
- Sheu SH, Liou CT, Tseng BC (1992) Optimal ordering policies and optimal number of minimal repair before replacement. *Microelectroni & Reliab* 32: 995–1002
- Silver EA (1981) Establishing reorder points in the (S,c,s) coordinated control system under compound Poisson demand. *Int J Prod Res* 19: 743–750
- Silver EA, Peterson R (1985) *Decision systems for inventory management*. Wiley, New York
- Silver EA, Pyke DF, Peterson R (1998) *Inventory Management and Production Planning and Scheduling*. John Wiley & Sons, New York
- Smith SA (1977) Optimal inventories for (S-1,S) system with no backorders. *Manag Sci* 23: 522–528
- Syntetos AA, Boylan JE (2005) The accuracy of intermittent demand estimates. *Int J Forecast* 21: 303–314
- Tango T (1978) Extended block replacement policy with used items. *J Appl Probab* 15: 560–572
- Tango T (1979) A modified block replacement policy using less reliable items. *IEEE Trans Reliab* 5: 400–401

- Taylor J, Jackson RRP (1954) An application of the birth-death process to the provision of spare machines. *Oper Res Q* 5: 95–108
- Teunter RH (2001) Economic ordering quantities for recoverable item inventory system. *Nav Res Logist* 48: 484–495
- Teunter RH (2004) Lot-sizing for inventory systems with product recovery. *Comput Ind Eng* 46: 431–441
- Thomas LC, Osaki S (1978) An optimal ordering policy for a spare unit with leadtime. *Eur J Oper Res* 2: 409–419
- Walker J (1997) Base stock level determination for insurance type spares. *Int J Qual & Reliab Manag* 14: 569–574
- Westerkamp TA (1998) Evaluating the maintenance process. *IIE Solut* 30(12): 22–27
- Willemain TR, Smart CN, Shockor JH, DeSautels PA (1994) Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *Int J Forecast* 10: 529–538
- Williams TM (1984) Stock control with sporadic and slow-moving demand. *J Oper Res Soc* 35: 939–948
- Xie M (1989) On the solution of renewal-type integral equations. *commun stat Theory & Method* 18: 281–293
- Yang H, Schrage L (2003) An inventory anomaly: risk-pooling may increase inventory. Tech Rep, University of Chicago
- Yang H, Schrage L (2009) Conditions that cause pooling to increase inventory. *Eur J Oper Res* 192: 837–851.
- Yurkiewicz J (2006) Forecasting software survey. *OR/MS Today* August 2006. <http://www.lionhrtpub.com/orms/surveys/FSS/FSS.html>. Accessed 12 April 2008.
- Zhang F, Jardine AKS (1998) Optimal maintenance models with minimal repair, periodic overhaul and complete renewal. *IIE Trans* 30: 1109–1119
- Zhang J (2005) Transshipment and its impact on supply chain members performance. *Manag Sci* 51: 1534–1539

Turnaround Maintenance

Salih Duffuaa and Mohamed Ben-Daya

10.1 Introduction

All major process industries (petrochemicals, refining, power generation, pulp and paper, steel plants, *etc.*) use Turnaround Maintenance (TAM) on a regular basis to increase equipment asset reliability, have continued production integrity, and reduce the risk of unscheduled outages or catastrophic failures. Plant turnarounds constitute the single largest identifiable maintenance expense. A major TAM is of short duration and high intensity in terms of work load. A 4 – 5 weeks TAM may consume an equivalent cost of a yearly maintenance budget. Because TAM projects are very expensive in terms of direct costs and lost production, they need to be planned and executed carefully. Turnaround management's potential for cost savings is dramatic, and it directly contributes to the company's bottom line profits. However, controlling turnaround costs and duration represent a definite challenge. Maintenance Planning and Scheduling is one of the most important elements in maintenance management and can play a key role in managing complex TAM events.

Turnaround maintenance (TAM) is a periodic maintenance in which plants are shut down to allow for inspections, repairs, replacements and overhauls that can be carried out only when the assets (plant facilities) are out of service. The overall objective of turnaround TAM is to maximize production capacity and ensure that equipment is reliable and safe to operate. Although different TAM may have different specific objectives, the following may constitute a list of the main objectives for TAM:

1. To improve efficiency and throughput of plant by suitable modification;
 2. To increase reliability/availability of equipment during operation;
 3. To make plant safe to operate till next TAM;
 4. To achieve the best quality of workmanship;
 5. To reduce routine maintenance costs;
 6. To upgrade technology by introducing modern equipment and techniques;
- and

7. To modify operating equipment to cope with legal requirements and or obligations such as environmental regulation.

During TAM, the following types of work are usually performed:

- Work on equipment which cannot be done unless the whole plant is shutdown;
- Work which can be done while equipment is in operation but requires a lengthy period of maintenance work and a large number of maintenance personnel;
- Defects that are pointed out during operation, but could not be repaired, will be maintained during turnaround period; and
- Upgrading equipment and introducing new technologies to improve speed and efficiency.

TAM originated in process industries and it plays an important role in maintaining consistent means of production delivered by reliable equipment. Because of the complexity and size of the TAM project in most process plants, the successful accomplishment of this event in terms of quality and cost is vital to the profitability of the company and to its competitive advantage. It has been shown by Joshi (2004) in a benchmarking study for more than 200 TAMs that most TAM experience schedule slips and costs overruns that is caused by inadequate planning and coordination. Therefore it is important for companies conducting TAM to have a sound process for planning, complete scope definition, a good strategy for execution that include integrated teams efforts. Lenahan (1999), Duffuaa *et al.* (1998) and Duffuaa and Ben Daya (2004) provide structured approaches for planning and managing TAM. The approaches in Lenahan (1999) and Duffuaa and Ben-Daya (2004) are in line with the guide of the project management body of knowledge provided by the Project Management Institute (2000).

Gupta and Paisie (1997) present a method for developing a TAM scope using reliability, availability and maintainability principles. They presented a team approach that utilizes the expertise and experience of the key personnel to analyze the plant's operational and maintenance data and economically justify every scope item. A risk-based method to optimize maintenance work scope is also presented by Merrick *et al.* (1999). The approach is similar to the ranking process used in failure modes and effects analysis. Krings (2001) stresses a proactive approach to shutdowns by allowing enough time for quality planning. Fiitipaldo (2000) reports the experience of the planning and execution of a desulfurization plant TAM. The report stresses the importance and need of the knowledge of competent, experienced employees and the use of basic quality assurance concepts throughout the process. Oliver (2002) discusses the TAM planning process and distinguishes TAM from other projects. The work process for planning a TAM must address the specific needs and challenges that are parts of repairing process equipment.

The purpose of this chapter is to outline a structured process of managing TAM projects. Sound procedures must be in place to make the process of conducting TAM more efficient and cost effective. The chapter covers all the phases of TAM from its initiation several months before the event till the

termination and writing of the final report. In particular, all aspects relating to the following phases of TAM are covered:

1. *Initiation*: this phase covers all strategic issues and activities needed to start the planning process. This includes TAM organization and compiling an initial work list.
2. *Preparation*: this is the critical phase of TAM. The successful execution of TAM hinges on excellent preparation. The most important activity in this phase is the determination of the work scope which is the basis of the whole planning process. This phase include preparation of the job packages, selection of contractors, defining safety, quality and communication programs. In addition to preparing the final budget for the project.
3. *Execution*: the phase is concerned with conducting the work, monitoring its progress and controlling various TAM activities so that the project is carried out on schedule and within budget.
4. *Termination*: this phase closes the project and assesses performance to document lessons learned, that may be used to improve future events.

The remainder of this chapter is organized as follows: Section 10.2 addresses TAM initiation followed by the topics related to work scope determination in Section 10.3. Section 10.4 deals with preparation of long lead time resources. Issues dealing with contractors, TAM planning and organization are discussed in Sections 10.5, 10.6, and 10.7, respectively. TAM site logistics and budget are presented in Sections 10.8 and 10.9 followed by the important aspects of quality and safety in Section 10.10. Sections 10.11 and 10.12 address TAM communication procedures and TAM execution. Finally the final report and its content are included in Section 10.13.

10.2 Turnaround Initiation

It is good practice to initiate TAM early enough to allow for forward and proper planning. The effective execution of TAM hinges on good planning. Without well-planned and executed shutdowns, equipment reliability suffers, and the plant pays the price of poor quality and lost production as demonstrated by previous studies, Duffuaa and Ben Daya (2004) and Joshi (2004).

It is necessary to form a TAM management team from experienced persons in the plant and it should include planners and engineers who are stakeholders and have the authority to make decisions concerning TAM. The complexity and size of the TAM project require a full time TAM manager who plays the key role in the TAM organization. His responsibility is to make sure that all activities are well planned and executed as planned.

In the initiation and preparation phases of TAM, detailed planning and preparation of all aspects of the project should be conducted. This includes the following items:

1. Work scope;
2. Job packages;
3. Pre-shutdown work;
4. Procurement of material and items;
5. Contract work packages and TAM contractors;
6. Integrated TAM plan;
7. TAM organization;
8. Site logistics plan;
9. TAM Budget;
10. Permission to work system;
11. Safety program;
12. Quality program;
13. Communication protocol;
14. Work control process;
15. Plant start-up procedure; and
16. TAM closing process.

The main elements of TAM and their descriptions are presented in the next sections of this chapter.

10.3 Work Scope

The work scope is the list of tasks or activities that need to be carried out during TAM. This is the foundation upon which all other aspects of the event revolve, especially safety, quality, duration, resource profile, and material and equipment requirements.

TAM work consists of a mix of maintenance tasks and project work. The lists of these activities can be divided into the following categories:

- Projects;
- Major maintenance tasks such as the overhaul of a large turbine or the re-traying of a large distillation column;
- Small maintenance tasks such as the cleaning and inspection of a small heat exchanger; and
- Bulk work such as the overhaul of a large number of small items such as valves and small pumps.

These activities are generated from various sources such as statutory safety requirements, production or quality improvement programs. Input about these lists is obtained from the production, maintenance, engineering, projects and safety departments.

A good strategy is to keep the TAM work list as short as possible commensurate with protecting the reliability of the plant. It is the job of the TAM teams and TAM manager to develop criteria for accepting work. Then process all work and project requests in a systematic way according to the set criteria and rules. This process is used to ensure that the approved work scope contains only

what is necessary to restore, maintain or enhance the reliability of the plant and which cannot be done at any other time.

Each work order should be planned before execution. Each planned job is accompanied by a work package, which is a written document containing all information needed to execute the work. It is very important that adequate personnel be dedicated to planning work packages.

The work package includes:

- A clearly defined scope of the work to be done;
- An estimate of the manpower required;
- A clear procedure and instructions for performing the work;
- A complete list of all tools and equipment needed;
- All non-standard tools acquired and staged at work site;
- A detailed list of spare parts required;
- All necessary permits;
- Drawings, sketches, special notes, and photographs, if needed;
- Contact information, should questions arise;
- Coordinated vendor support *etc.*;
- Schedule for execution for each type of craft;
- Safety and environmental hazard precautions; and
- Personal Protective Equipment needed.

Any work that is placed on the shutdown schedule that is not fully planned will effectively place the burden of planning on the people doing the work. This creates confusion, causes delays, and creates opportunities for mistakes and hazards. It is always much safer to execute planned work, since possible hazards are systematically identified and avoided.

10.4 Long Lead Time Resources

The procurement of material and spare parts should be done in advance, especially items that require long lead time. For example, the delivery time for a compressor rotor might well be as much as 16 months. In order to identify these, it is necessary to analyze the work list as early as possible to ensure that sufficient time is allowed for ordering these items. Special attention should be given to the following activities:

- Pre-fabricated work;
- Special technologies that are needed;
- Vendors' representatives; and
- Services and utilities.

10.5 Contractors

The size and complexity of TAM necessitate the use of contractors. A good preparation of the job packages is essential in determining the type and amount of contracting required. The size of the workforce needed in some circumstances exceeds 15 times the size of the regular in house maintenance personnel. Having a well prepared work-packages provide the company a leverage when negotiating with contractors. Other reasons for using contractors include:

- Special skills;
- Experience and professionalism;
- Productivity, cost and efficiency; and
- Controlling the duration of TAM.

It is important to have a good process for contract selection to ensure the delivery of quality work. It is sometime a good practice to have one experienced person from the plant plan supervising 10–15 contractors' personnel.

10.6 TAM Planning

TAM requires effective detailed planning due to the fact that it involves a large number of personnel working under time constraints to accomplish a lot of work. It therefore requires planning of an order of detail that is not found during normal operation. The basic objective of planning is to ensure that the right job is done at the right time and assigned to the right people.

Planning of a TAM requires the participation of and active co-operation of several integrated teams organizing and delivering many projects and jobs. Table 10.1 shows the responsibilities and roles of these teams.

Table 10.1. People involved in TAM planning and their role

Team	Role
Preparation team	Prepare the master plan
Plant team	Provide basic data, work requests, technical information and the shutdown-startup network and then validate final planning documentation
Inspectors	Specify inspection work, requirements and techniques
Engineering	Provide technical information and support
Project managers and engineers	Provide the planning and documentation for project work
Contractors representatives	Advise on feasibility of their part of the plan
Policy team	Approve and fund the final plan

It is good practice to categories TAM work into classes based on sound criteria. Common practice is to put them in three categories: major tasks, minor tasks, and bulk work.

TAM planning also covers the following important aspects:

- The shutdown startup logic;
- The shutdown network;
- The startup network;
- The critical path program; and
- Work scheduling.

The final TAM schedule will be an optimized blend of the above requirements. Work schedules are usually generated using project management techniques and software.

It is not uncommon to discover additional work when the execution of TAM starts and therefore it is wise to have contingency plans and flexibility built in a TAM schedule.

10.7 TAM Organization

TAM organization is critical to the success of the event and deals with addressing the following two important questions:

1. Who will manage the turnaround? and
2. Who will carry out the work?

The plant management must utilize previous experience and select the most suitable personnel to plan and execute TAM. A number of basic principles have been developed out of past experience and can help in TAM organization:

- A turnaround is a task oriented event;
- The minimum number of people should be used;
- The TAM organization is hierarchic;
- One person must be in overall control;
- Single point responsibility is exercised at every stage;
- Every task is controlled at every stage; and
- The organization is a blend of the required knowledge and experience.

A good organization would blend the following:

- *Plant personnel*, who possess local knowledge;
- *TAM personnel*, skilled in planning, coordination and work management;
- *Technical personnel*, who possess engineering design and project skills; and
- *Contractors*, and others who possess the skills and knowledge to execute the work.

10.8 Site Logistics

The preparation of well documented site logistics facilitates the execution of TAM, minimizes delays and improves utilization of resources. The site logistics organizes TAM operations and shows places of storage of material, location of equipment, accommodation of contractor's personnel and the effective mobilization of every one to perform prescribed tasks. An important element in the site logistic is the plot plan. The list of elements that are usually shown in plot plan includes but not limited to the following:

- Plant perimeter and boundaries;
- All major items of plant equipment and pipe-work;
- All roads;
- Areas or roads where access is prohibited;
- All accesses to site and site roads;
- All locations of fire assembly locations;
- Layout areas for contaminated material;
- Areas designated for TAM storage and quarantine compounds;
- Areas for hazard substances;
- Approved vehicle routes with direction of traffic;
- Location of additional safety equipment;
- Temporary piping and cabling for utilities;
- Areas for various technical work such welding, air compressors, *etc.*;
- Various contractors areas;
- Parking areas; and
- Sites for TAM control, induction and safety.

10.9 TAM Budget

Experience has shown that TAM events are expensive and they usually experience cost overruns. It is important to have a good and accurate cost estimate for the TAM projects and jobs. Well planned job packages are a prerequisite for good costs estimates.

The main elements of the budget are the following:

- TAM planning and management;
- Company labor;
- Contractors;
- Spares and materials;
- Equipment purchase and rent;
- Accommodation facilities;
- Utilities; and
- Contingencies.

The TAM policy team must approve the final cost estimates. If the estimate is more than the allotted budget, several options can be explored in order to secure the deficit or bring the cost back within the budget figures. This can be done by eliminating and/or deferring some tasks that do not compromise the integrity of the plant. After the budget is approved it is the responsibility of TAM team to control costs and minimize costs over runs.

10.10 Quality and Safety Plans

10.10.1 Quality Plan

The quality plan is a process by which TAM teams ensure that the tasks are performed according to standards. It also ensures that the quality of jobs are planned, executed and controlled. The quality plan includes:

- A quality policy that is a statement that guides the practices and behaviors essential for high quality work performance throughout the plant during TAM. The policy has to be properly and consistently implemented by all concerned.
- A system to ensure the implementation of the quality policy.
- A quality plan for each critical job that may affect plant reliability or integrity.

To assure quality, the quality plan should ensure that the requirements of every task must be correctly specified and then performed to that specification. The way to get it right is to have a coherent, auditable quality trail from initial work request to final acceptance of the completed task.

10.10.2 Safety Plan

TAM brings a large number of people into a confined area to work under pressure of time with hazardous equipment. The targets set for safety must be high – zero accidents, incidents, fires, *etc.* To meet the safety targets a well designed safety plan is needed. A safety plan includes:

- A safety policy which is a statement that guides the practices and behaviors essential for high quality safety performance throughout the plant during TAM. The policy must be well documented and communicated in order to facilitate its implementation.
- Safety communication network that establishes the hierarchy responsible for setting the safety policy and ensuring that everyone adheres to it. It specifies the safety chain and clear lines of communication. The safety chain consists of TAM manager, engineers, supervisors and workers.
- Safety working routine to ensure that the necessary steps are taken to eliminate the hazards and to protect workers against them. The working routine consists of the following elements:

- Work permit;
- Work environment;
- The worker;
- The task specification;
- Material and substances; and
- Tools and equipment.

The company must have a scheme for monitoring safety performance. The scheme must include daily inspection, spot checks on specific jobs and a program to involve everyone in safety monitoring. The criteria focus on four main factors. These factors are:

- Daily safety theme awareness;
- Unsafe acts;
- Unsafe conditions; and
- House keeping.

The program that involves everyone in monitoring safety depends on the organization culture and safety awareness programs. One program could be to empower everyone to report unsafe acts or conditions. Everyone who reports a genuine unsafe act or situation must be rewarded. Also senior management should be involved and make daily safety tours and distribute the daily rewards and awards for the best safety practice. An alternative way is to establish a daily newsletter and awards for the safety man of the day. Criteria must be developed to select the safety man of the day.

10.11 TAM Communication Procedures

The execution of TAM gathers many persons from different contractors. Most of the persons participating in the TAM event are new to the site and come from organizations with different cultures and attitudes towards safety and quality. The diversity of people involved could give rise to conflict and competition. In such an environment communication plays a central role in reducing delays, conflict and accidents. The communication plan must specify the following:

- What to communicate?
- Whom to communicate to?
- When to communicate?
- Who should do the communication? and
- How to communication in the most effective way?

The following three briefings need to be an integral part of a TAM communication package:

- The general briefing;
- The major tasks briefing; and
- Daily briefing or reporting.

10.12 TAM Execution

Execution of TAM starts after careful planning, preparation and ensuring all the required resources are on site. The focus at this stage is delivering the jobs and projects as planned through effective monitoring and control. The execution process involves several key steps that include:

1. The plan is finalized and the event schedule is complete;
2. A plan for the unexpected work is also devised;
3. TAM manager routine;
4. The day shift;
5. Shift change procedure;
6. Night shift;
7. In control of work;
8. Control of cost;
9. Daily program;
10. Daily reporting; and
11. Start of the plant.

Other important issues during execution include:

- Guidelines for plant shutdown;
- Sample of daily routine;
- Work control guidelines;
- Handling of unexpected work; and
- Start-up guidelines.

10.13 TAM Closing and Final Report

After completing TAM tasks, the plant start-up procedure will follow. The plant operation is not the end of TAM. It is necessary to review the whole event to gather and document lessons learned. This necessitates the preparation of TAM final report.

The content of TAM final report should address the following important topics:

1. TAM policy;
2. The work scope;
3. The preparation phase;
4. Planning;
5. The organization;
6. Control of work;
7. Contractor performance;
8. Safety;
9. Quality;
10. Site logistics;
11. Communications; and
12. Recommendations for improvement.

The organization must measure TAM performance and observe trends. As with all measurements, a single indicator can mislead. It is therefore necessary to design a number of criteria to provide a balanced indication of performance. Having a work process does not guarantee a successful TAM, but benchmarking and continuous learning from previous events considerably reduces the likelihood of failure.

10.14 Conclusion

We conclude this chapter by highlighting the main differences between TAM projects and other projects and some lessons learned from a survey conducted by the authors in the petrochemical industries. TAM management is project management in the sense that it has all the elements of project management. In general the project management body of knowledge applies to TAM. However, there are unique features that distinguish TAM from other projects. The main differences are as follows (Lenahan, 1999):

1. Most projects create something new, as in the constructions industry. However TAM management deals with the replacement, repair or overhaul of equipment.
2. The work scope of other projects is well defined and visible based on drawings, specification, contracts, *etc.*, while TAM scope is loosely defined and need to be identified based on past TAM experience, inspection reports, operations requests, *etc.*
3. In regular projects uncertainties are usually imposed by the operating environment, such as delivery of materials, availability of labor, the weather, *etc.* In TAM, the degree of wear and damage is unknown until the plant is opened for inspection, which an additional sources of uncertainty that is difficult to control.
4. TAMs are usually duration driven, which may require the mobilization of hundreds and sometimes thousands of workers that need to complete a lot of work in a short duration.

A survey has been conducted in the petrochemical industry (Duffuaa *et al.* 1998) which found that the surveyed plants are doing very good job in planning and executing TAM. They are using acquired experience to minimize the duration of TAM over the years through effective planning and control. The following aspects of TAM are performed well:

- Allowing a good planning horizon for TAM;
- Preparing job work packages;
- Contractors' selection criteria;
- Ensuring safety during TAM;
- Communication and reporting during TAM period; and
- Execution and control of TAM.

Elements that industry must pay attention to and which constitute essential directions for continuous improvement and success in conducting TAM are:

- Documentation of major TAM processes and procedures;
- Establishment of TAM maintenance manual;
- Developing TAM performance measurements;
- Standardizing TAM final report;
- Strengthen the process of feedback and learning from previous TAM experiences;
- Integration of TAM with existing maintenance management information system (MMIS); and
- Costing and cost reduction in TAM.

References

- Duffuaa SO, Ben-Daya M (2004) Turnaround Maintenance in Petrochemical Industry: Practices and Improvement. *J of Qual in Maint Eng* 10: 184–190.
- Duffuaa SO, Raouf A and Campbell JD (1998) Planning and Control of Maintenance Systems: Modeling and Analysis, Wiley, New York
- English P (1997) Planning, Scheduling, Execution Smooth Maintenance Shutdowns. *Pulp and Pap*, July: 93–96
- Fiitipaldo JJ (2000) Desulfurization Plant Turnaround: Planning and Execution. *AISE Steel Technology*, April: 42–46
- Gupta SK and Paisie JE (1997) Reduce Turnaround Costs. *Hydrocarb Process*, Jan.: 67–74
- Joshi N (2004) Benchmarking and best practices for effective turnarounds, a working paper Independent Project Analysis, Inc. (IPA).
- Krings D (2001) Proactive Approach to Shutdowns Reduces Potlatch Maintenance Cost, *Pulp and Pap*, December: 27–29
- Lenahan T (1999) Turnaround mangement, Butterworth-Heinemann
- Merrick E, Leonard R, Eckhardt P, and Baughman H (1999) Risk-Based Methods Optimize Maintenance Work Scope, *Oil & Gas Aug*: 47–52.
- Oliver R (2002) Complete Planning for Maintenance Turnarounds will Ensure Success, *Oil & Gas J* April: 54–62.
- Project Management Institute (2000) A guide to the Project Management Body of Knowledge, 2000 Editions. <http://www.pmi.org/>

Maintenance Planning and Scheduling

Maintenance Planning and Scheduling

Umar M. Al-Turki

11.1 Introduction

Planning is the process of determining future decisions and actions necessary to accomplish intended goals, and targets. Planning for future actions helps in achieving goals in the most efficient and effective manner. It minimizes costs and reduces risks and missing opportunities. It can also increase the competitive edge of the organization. The planning process can be divided into three basic levels depending on the planning horizon:

1. Long range planning (covers a period of several years);
2. Medium range planning (one month to one year plans); and
3. Short range planning (daily and weekly plans).

Planning is done at different decision levels, strategic or tactic. It can be done at different organizational levels, corporate, business, functional or operational. Decision at the strategic level are concerned with issues related to the nature of existence of the business as a corporate whereas tactical decisions effect the way business conducted at a certain stage of its growth line. Strategic planning sets the long term vision of the organization and draws the strategic path for achieving that intended vision. Long term and short term planning at the tactical level is concerned on selecting ways within a preset strategy for achieving long, medium and short term goals and targets. Strategic planning is by definition a long term plan and can be done at the functional, business or corporate level. Long term planning, however, is not necessary strategic. In general, regardless of the type and purpose of planning, it includes the determination of the actions or tasks as well as the resources needed for their implementation.

Scheduling is the process of putting the tasks determined by the plan into a time frame. It takes into consideration the intended goals, the interrelations between the different planned tasks, the availability of resources overtime and any other internal and external limitations and constraints. The quality of the resulting schedule is usually measured by a performance measure in relation to the intended

goal of the task or tasks. Performance measures can be related to different types of costs through meeting due dates, time of completion, or utilization of resources.

Maintenance in its narrow meaning includes all activities related to maintaining a certain level of availability and reliability of the system and its components and its ability to perform at a standard level of quality. It includes activities related to maintaining spare part inventory, human resources and risk management. In a broader sense, it includes all decisions at all levels of the organization related to acquiring and maintaining high level of availability and reliability of its assets. Maintenance is becoming a critical functional area in most types of organizations and systems such as construction, manufacturing, transportation, *etc.* It is becoming a major functional area that effects and affected by many other functional areas in all types of organizations such as production, quality, inventory, marketing and human resources. It is also getting to be considered as an essential part of the business supply chain at a global level. This increasing rule of maintenance is reflected in its high cost which is estimated to be around 30% of the total running cost of modern manufacturing and construction businesses. A system view of a maintenance system is introduce by Visser (1998) that puts maintenance in perspective with respect to the enterprise system as shown in Figure 11.1.

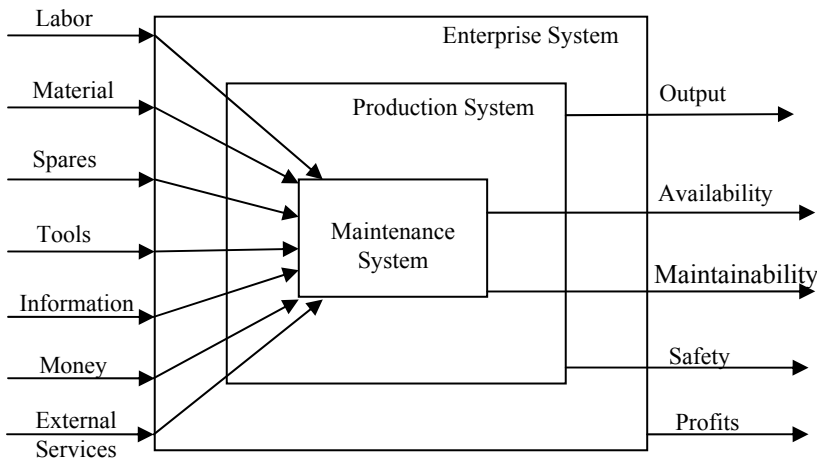


Figure 11.1. Input output model of the enterprise

Corporate business planning, long or short term, strategic or tactic should take maintenance into consideration for all types of decisions that involve future major investments. A decision on acquiring a new facility, for example, might turn into a complete disaster for the whole business for its low maintainability. Capacity planning of the plant should consider its maintainability and the capacity of maintaining it.

Planning and scheduling are the most important aspects of sound maintenance management. Effective planning and scheduling contributes significantly to reducing maintenance costs, reducing delays and interruptions and improving

quality of maintenance work by adopting the best methods, procedures and assigning the most qualified crafts for the job. The principal objectives of maintenance planning and scheduling include:

- Minimizing the idle time of maintenance forces;
- Maximizing the efficient use of work time, material, and equipment; and
- Maintaining the operating equipment at a level that is responsive to the need of production in terms of delivery schedule and quality.

Maintenance as a major function in the organization should have its own strategic plan that aligns its objectives and goals with the objective and goals of the whole organization. Strategies for maintenance operations should be selected among alternatives to achieve these objectives. Outsourcing is one of the common strategies in many business environments that are usually used as alternative strategy for building the maintenance capacity internally. Few papers have been written lately that discusses strategic maintenance planning including Tsang (1998, 2002) and Murthy *et al.* (2002). An alternative strategy combines the first two alternatives in different forms including outsourcing some maintenance functions and self maintaining some other critical function. A discussion of advantages and disadvantage of outsourcing is discussed by Murthy *et al.* (2002).

Any planning activity at any level should start by forecasting the future at that level. Strategic level forecasting is concerned with future trends and possible changes in the business itself or in its environment in the long run. Long term forecasting is mainly concerned with the future demand of its outcomes in the long range which is usually a year or a few years. Middle term forecasting focuses on demand on a monthly basis for 1 year. Different forecasting techniques are available for different types of forecasting varying between highly qualitative for long term forecasting to highly quantitative for middle and short term forecasting. Forecasting will not be discussed in this chapter in detail since it is part of another chapter in this handbook.

Planning maintenance operations under clear maintenance strategies and strategic objectives sets the direction for middle and short term maintenance planning. Having the appropriate future forecast, plans are developed, in line with the developed strategies, to achieve the intended goals of the maintenance operations which usually supports the overall goals of the business unit in the short, medium or long term. As a result a set of decisions and actions are set to meet the expected forecast at the right time in the optimum manner with respect to the overall goal of the organization. These decisions are usually related to resource availability such as human resources in quantity and quality (skills), tools and equipment. Varieties of quantitative techniques are available to support the planning process in the medium and short range such as mathematical modeling and simulation.

Short term planning is usually followed by scheduling which is the process of putting the planned activities in their time frame in relation to each other. Usually scheduling is coupled with short term planned activities. These activities are scheduled for implementation on the available (planned) resources so that a certain objective is achieved. Having a set of planned maintenance activities to be conducted at a certain period of time (a week for example) the scheduling process

is concerned with allocating the right maintenance crew and the right equipment at the time that satisfies the intended requirement in terms of time and quality. Having limited resources and unplanned activities makes the scheduling task extremely complicated. Quantitative tools are designed to assist the scheduler in building the most efficient schedules that are robust to changes in the environment.

The objective of this chapter is to give hands on knowledge of maintenance planning and scheduling for planners and schedulers at all levels. This knowledge will help in the development of the most effective and efficient plans and schedules of maintenance operations. Planners for maintenance at the corporate level are introduced, in the next section, for different dimensions and options of strategic maintenance planning. Each dimension, including outsourcing and contractual relationships, organization and work structure, maintenance methodology and supporting systems, is discussed for the risks and benefits of each possible option. Middle level planners are usually concerned with medium range maintenance planning which is introduced in Section 11.3 with its components and steps for sound development. Lower level planners concerned with short range plans are addressed in Section 11.4. Middle level maintenance planners, as well as short level planners, are usually involved in scheduling activities and tasks over their concerned time range. Elements of maintenance scheduling are introduced in Section 11.5 followed by scheduling techniques in Section 11.6. Section 11.7 highlights some aspects of information system support available for maintenance planning and scheduling that is usually a concern of strategic level planners and utilized by planners and schedulers at all levels.

11.2 Strategic Planning in Maintenance

Traditionally, maintenance is not viewed as a strategic unit in the organization and hence maintenance planning was mostly done at midterm range. However, the strategic dimension of the maintenance function has lately drawn the attention of the researchers and practitioners with the increase in the competition at a global level and with the increase of the maintenance cost relative to other costs in the organization. Equipment availability, especially in certain business sectors like energy generation and oil exploration and other mega projects, is becoming a major concern because of its high cost of acquisition. Emerging operational strategies such as lean manufacturing are shifting the emphasis from volume production to quick response, defect prevention and waste elimination. These changes in operations strategies require changes in maintenance strategies related to equipment and facility selection and optimizing the maintenance activities with respect to the new operations objectives. Rapid technological changes in non-destructive testing, transducers, vibration measurement, thermography, and other emerging technologies generated an alternative strategy of condition based maintenance. However, these new technologies introduced new challenges that maintenance systems have to face including the development of new capabilities and management practices to utilize these technologies. Plans have to be developed at a strategic level for keeping up with emerging technologies in the long run.

These changes in the business environment developed the realization that maintenance must not be viewed only in the narrow operational context dealing

with equipment failure and their consequences. Rather it must be viewed in the long term strategic planning context that integrates technical and commercial issues as well as changes in the sociopolitical trends. Maintenance must be viewed strategically from the overall business prospective and has to be handled within a multidisciplinary approach. This approach takes into consideration the sociopolitical, demographic trends and the capital needed. See Murthy *et al.* (2002). It deals with strategic issues such as outsourcing of maintenance and the associated risks and other related issues.

Murthy *et al.* (2002) describes the strategic view of maintenance by the equipment state, the operating load, maintenance actions (strategies) and business objectives. The state of the equipment is affected by the operating load as well as the maintenance actions. The operating load is dependent on the production plans and decisions which are in turn effected by commercial needs and market consideration. Therefore, maintenance planning has to take into consideration the production planning, maintenance decisions, equipment inherited reliability and market and commercial requirements. The model is shown in Figure 11.2.

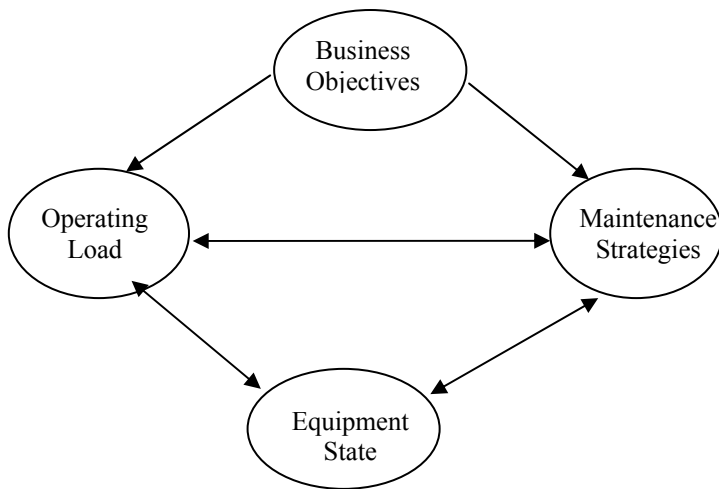


Figure 11.2. Key elements of strategic maintenance management

Four strategic dimensions of maintenance are identified by Tsang (2002) in relation to the system view of the enterprise shown earlier in Figure 11.1.

The first dimension is the service delivery strategy. Outsourcing *vs* in-house maintenance are two possible alternatives for maintenance delivery strategies. Many petrochemical processing plants outsource all their equipment and facility maintenance. Others outsource particular specialized or risky aspects of maintenance. A survey conducted in North America, and cited by Campbell (1995), found that 35% of companies surveyed outsource some of their maintenance. The potential benefits of outsourcing maintenance activities include less hassle, reduced total system costs, better and faster work done, exposure to

outside specialists, greater flexibility to adopt new technologies and more focus on strategic asset management issues (Watson, 1998; Campbell, 1995).

The selection between the two options should not be regarded as a tactical matter; instead it should be made in the context of the overall business strategy. Murthy *et al.* (2002) had explored the two alternatives and discussed the long term costs and risks of each alternative. Some general guidelines are laid out in relation to this issue including that maintenance management and planning should not be outsourced; the maintenance implementation, however, may be outsourced based on cost and risk consideration. Risks are very much linked to the service supply market. Having a single dominating supplier in the market makes the user company hostage to that supplier services. On the other hand, if the suppliers are weak they might not be able to supply quality and reliable service as much as the internal service can. Lastly, the service should not be outsourced if the company does not have the capability to assess or monitor the provided service and when it lacks the expertise in negotiating sound contracts. Contracts have to be written carefully to avoid long term escalation in its costs and risks. Tsang (2002) have an excellent analysis of the two options in terms of things that should not be outsourced. An activity that is considered to be the organization's core competency should not be outsourced. An activity may be considered as a core competency if it has a high impact on what customers perceive as the most important service attribute or the activity that requires highly specialized knowledge and skills. The costs involved in the internal service include personnel development and infrastructure investment and managing overhead. The costs involved in the outsourcing include the costs of searching, contracting, controlling and monitoring.

Contractual relationship with the service provider is an important aspect of outsourcing. The benefits of outsourcing are seldom realized because of contracts that are task oriented rather than performance focused and the relationship between the service provider and the user is adversarial rather than partnering. In the absence of long term partnership between maintenance service supplier and the user, the supplier will be hesitant to invest in staff development, equipment and new technologies. The relationship between the supplier and the user is determined by the type of contract. See Martin (1997) for different types of contracts.

While outsourcing has great potential for significant benefits, it also includes some potential risks such the following:

- Loss of critical skills;
- Loss of cross functional communications; and
- Loss of control over a supplier.

To reduce the risks, the contract and the contracting process should be dealt with in a delegated manner. Specialists in the maintenance technical requirements and specialists in technology and business needs as well as specialist in contract management should be involved in the process. The contract itself should have a conflict resolution and problem solution mechanism for uncertainties and inevitable changes in the requirements and technology changes. Other measures for reducing risks include splitting maintenance requirements to more than one supplier.

The second dimension of strategic maintenance management identified by Tsang (2002) is the organization and work structure. Traditionally, the organization structure is hierarchical and highly functionalized within which maintenance is organized into highly specialized trades. This organization has led to many problems in terms of efficiency and effectiveness. New process oriented organization structures are emerging for more effective and efficient management of business units. Within these structures, maintenance is viewed as part of a group owning the process. Different work structures may be considered for different types of maintenance work. Choices between plant flexible and plant specialized tradesman, centralized vs dispersed workshops, trade specialized vs multi-skilled trade-force have to be made.

The third dimension of strategic maintenance management is the maintenance methodology. There are four basic approaches to maintenance: run to failure, preventive maintenance, condition based maintenance and design improvement. Methodologies for selecting the most suited approach such as reliability-centered maintenance and total productive maintenance are developed and adopted by many companies around the globe. The choice between these methodologies is a strategic decision that has to be made based on the organizations global objectives. The details of these issues are introduced and discussed in other chapters of this handbook.

The fourth dimension of strategic maintenance management is the selection of the support system that includes information system, training, and performance management and reward system. Each element has to be carefully selected to support the overall objective of the organization. Enterprise Resources Planning, ERP, systems are gaining ground in large organizations and to a certain extent in medium size organizations. The power of ERP lies in its ability to integrate different functional areas within the organization which is an essential requirement for maintenance planning and scheduling. Successful implementation of the system requires careful system selection and implementation strategy that is human focused. For details about integrating maintenance strategies in ERP see Nikolopoulos *et al.* (2003).

In summary, the maintenance strategy is developed based on the corporate objectives and in line with its strategies. The maintenance strategy is based on a clear vision of the role maintenance playing in the corporate strategy and on clear objectives that are in line with the corporate objectives. Strategic choices have to be made in relation to organization structure, maintenance methodologies, supporting systems and outsourcing related decisions. Once selections are made, middle range plans have to be made regarding capacity and workforce planning. Weekly and daily plans are then made and activities are scheduled for implementation followed by measuring performance for continuous feed back for improvement. The maintenance planning process is summarized in the model shown in Figure 11.3 and discussed in details in the remainder of the section.

There are different alternative methodologies for the strategic planning process. All of them stress the involvement of all stakeholders in the process through brain storming sessions and focused group meetings. One possible methodology comprise of the following steps:

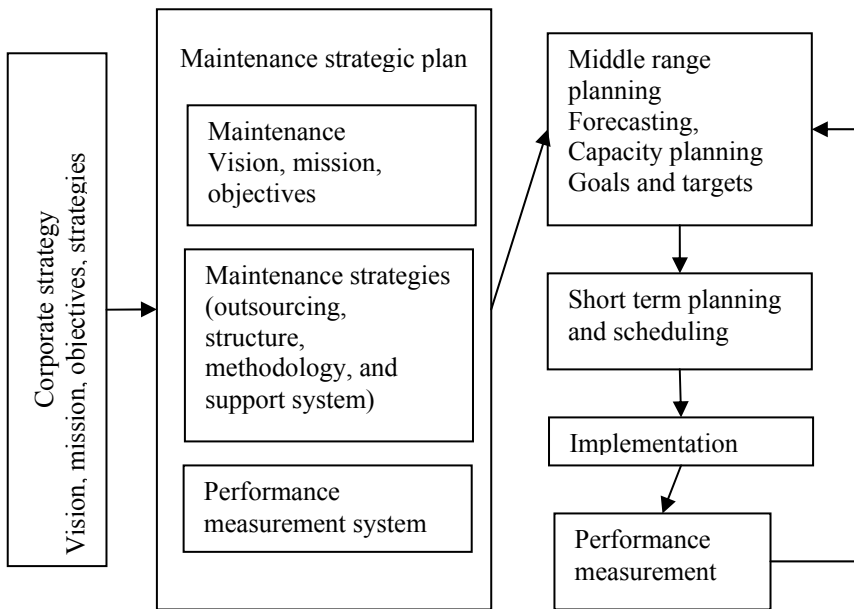


Figure 11.3. The maintenance planning process

1. Revise the corporate vision, mission and objectives and identify the rule of maintenance in achieving them.
2. Formulate the identified rule as a mission statement for maintenance.
3. Set the strategic objectives of maintenance.
4. Develop a set of quantitative measures for the identified objectives.
5. Evaluate the current situation in terms of achieved objectives and identify the gap between the actual and the desired situation.
6. Analyze the current internal and external situation related to the maintenance function. A common methodology is conducting SWOT analysis (identify internal strengths and weaknesses and external opportunities and threats).
7. Select a strategy for each of the four dimensions discussed in this chapter that would achieve the objectives in the most efficient and effective manner based on the gap identified in step 5 and the situation analysis conducted in step 6.
8. Develop a system for continuous situation assessment and strategic adjustment.

11.3 Medium Range Planning

The medium range plan covers a period of 1 month to 1 year. The plan specifies how the maintenance force operates and provides details for major overhauls, construction jobs, preventive maintenance plans, plant shutdowns, and vacation

planning. A medium range plan balances the need for manpower over the period covered and estimates the required spare parts and material acquisition. Medium range planning needs utilization of the following methods:

1. Sound forecasting techniques to estimate the maintenance load;
2. Reliable job standard times to estimate manpower requirements; and
3. Aggregate planning tools such as linear programming to determine optimum resource requirements.

For planning purposes maintenance work can be classified into the following five categories:

1. *Routine and preventive maintenance*, which includes periodic maintenance such as lubricating machines, inspections and minor repetitive jobs. This type of work is planned and scheduled in advance.
2. *Corrective maintenance*, which involves the determination of the causes of repeated breakdowns and eliminating the cause by design modification;
3. *Emergency or breakdown maintenance* is the process of repairing as soon as possible following a reported failure. Maintenance schedules are interrupted to repair emergency breakdowns.
4. *Scheduled overhaul*, which involves a planned shutdown of the plant to minimize unplanned shutdowns.
5. *Scheduled overhaul*, which involves repairs or building of equipment which does not fall under the above categories.

The maintenance management system should aim to have over 90% of the maintenance work to be planned and scheduled in order to reap the benefits of planning and scheduling.

Maintenance planning and scheduling methodologies and techniques are developed in line with production planning methodologies as it is viewed as a special type of production system. However, the two systems differ in several aspects:

1. The demand for maintenance work has more variability than production and the arrival of the demand is stochastic in nature.
2. Maintenance jobs have more variability between them, even the same types of jobs differ greatly in content. This makes job standards hard to develop compared to production jobs. Reliable job standards are necessary for sound planning and scheduling.
3. Maintenance planning requires the coordination with other functional units in the organization such as, material, operations, engineering and in many situations it is a major cause of delays and bottlenecks.

The above reasons necessitate a different treatment for maintenance planning and scheduling.

Forecasting the maintenance work required of each type to keep a certain level of a predetermined objective is the most important step in the planning process. There is no best method of forecasting; instead a mixture of techniques is most

appropriate for highly stochastic interrelated type work as it is the case maintenance. Mixture of qualitative and quantitative techniques is usually used in forecasting the medium range need of maintenance work volume. Usually the maintenance work volume varies over time throughout the year for internal causes such as production volume and external causes such as weather conditions. However, planned maintenance activities can be used to smooth out the requirement over the planning period. Forecasting is discussed in detail in another chapter in this handbook.

Once the work volume is forecasted, it can be easily translated to workforce and tools and equipment requirements under the selected strategy regarding service delivery and maintenance methodology. Optimization techniques such as mathematical programming can assist planners in determining the most efficient and the least cost plan of maintenance workforce. It includes decisions related to temporary or permanent outsourcing for some of the maintenance work throughout the planning horizon.

The planning process comprises all the functions related to the preparation of the work order, bill of material, purchase requisition, necessary drawings, labor planning sheet, job standards and all the data needed prior to scheduling and releasing the work order. Therefore, an effective planning procedure should include the following steps as identified by Duffuaa (1999):

1. Determine job content (may require site visits).
2. Develop work plan. This entails the sequence of activities in the job and establishing the best methods and procedures to accomplish the job.
3. Establish crew size for the job.
4. Plan and order parts and material.
5. Check if special equipment and tools are needed and obtain them;
6. Assign workers with the appropriate craft skill.
7. Review safety procedures.
8. Set priorities (emergency, urgent, routine, and scheduled) for all maintenance work.
9. Assign cost accounts.
10. Fill the work order.
11. Review backlog and develop plans for controlling it.
12. Predict the maintenance load using an effective forecasting technique.

The medium range planning process is coupled with a scheduling process which is considered long range scheduling known as the master schedule. It is based upon the existing maintenance work orders including the blanket orders issued for routine and preventive maintenance, overhaul and shutdowns. It will reveal when it is necessary to add to the maintenance work or subcontract a portion of the maintenance work. The reliability of the master schedule depends heavily on the reliability of the forecast of maintenance work and the validity of the standard times and a reliable mechanism for controlling and recording maintenance activities. Nevertheless, the master schedule can be revised regularly to accommodate changes in the plan and more accurate information availability.

11.4 Short Range Planning

Short range planning concerns periods of 1 day to 1 week. It focuses on the determination of all elements required to perform industrial tasks in advance. Short range planning in the context of maintenance means the process by which all the elements required to perform a task are determined and prepared prior to starting the execution of the job.

The maintenance work order does not usually provide enough space to perform the details of planning for extensive repairs, overhauls or large maintenance projects. In such cases where the maintenance job (project) is large and requires more than 20 h, it is useful to fill a maintenance planning sheet. An example of such a sheet is given in Figure 11.4. Maintenance planning sheets were found useful in planning the maintenance of freight cars in railways, when the cars arrive for their six month scheduled preventive maintenance. In the maintenance planning sheet the work is broken down into elements. For each element the crew size and the standard times are determined. Then, the content of the planning sheet is transferred in one or several work orders. In filling the planning sheet or the work order the planner must utilize all the expertise available in the maintenance department. Thus consultations with supervisors, foremen, plant engineers and crafts should be available and very well coordinated.

Therefore the planning and scheduling job requires a person with the following qualifications:

- Full familiarity with production methods used through the plant;
- Sufficient experience to enable him to estimate labor, material and equipment needed to fill the work order;
- Excellent communication skills;
- Familiarity with planning and scheduling tools; and
- Preferably, with some technical education.

The planner office should be centrally located and the office organization depends on organization size.

11.5 Maintenance Scheduling

Maintenance scheduling is the process by which jobs are matched with resources (crafts) and sequenced to be executed at certain points in time. The maintenance schedule can be prepared in three levels depending on the horizon of the schedule. The levels are: (1) medium range or master schedule to cover a period of 3 months to 1 year; (2) weekly schedule, it is the maintenance work that covers a week; and (3) the daily schedule covering the work to be completed each day.

11.5.1 Elements of Sound Scheduling

Planning maintenance work is a prerequisite for sound scheduling. In all types of maintenance work the following are necessary requirements for effective scheduling:

1. Written work orders that are derived from a well conceived planning process. The work orders should explain precisely the work to be done, the methods to be followed, the crafts needed, spare parts needed and priority.
2. Time standards that are based on work measurement techniques;
3. Information about craft availability for each shift.
4. Stocks of spare parts and information on restocking.
5. Information on the availability of special equipment and tools necessary for maintenance work.
6. Access to the plant production schedule and knowledge about when the facilities may be available for service without interrupting the production schedule.
7. Well-defined priorities for the maintenance work. These priorities must be developed through close coordination between maintenance and production.
8. Information about jobs already scheduled that are behind schedule (backlogs).

The scheduling procedure should include the following steps as outlined by Hartman:

1. Sort backlog work orders by crafts;
2. Arrange orders by priority;
3. Compile a list of completed and carry-over jobs;
4. Consider job duration, location, travel distance, and possibility of combining jobs in the same area;
5. Schedule multi-craft jobs to start at the beginning of every shift;
6. Issue a daily schedule (except for project and construction work); and
7. Have a supervisor make work assignments (perform dispatching).

The above elements provide the scheduler with the requirements and the procedure for developing a maintenance schedule. Next, the role of priority in maintenance scheduling is presented together with a methodology for developing the jobs priorities.

11.5.2 Maintenance Job Priority System

The maintenance job priority system has a tremendous impact on maintenance scheduling. Priorities are established to ensure that the most critical and needed work is scheduled first. The development of a priority system should be well coordinated with operations staffs who commonly assign a higher priority to

maintenance work than warranted. This tendency puts stress on the maintenance resources and might lead to less than optimal utilization of resources. Also, the priority system should be dynamic and must be updated periodically to reflect changes in operation or maintenance strategies. Priority systems typically include three to ten levels of priority. Most organizations adopt four or three level priorities. Table 11.1 provides classification of the priority level and candidate jobs to be in each class as identified by Duffuaa *et al.* (1999).

Table 11.1. Priorities of maintenance work

Code	Name	Time frame work should start	Type of work
1	Emergency	Work should start immediately	Work that has an immediate effect on safety, environment, quality, or will shut down the operation
2	Urgent	Work should start within 24 h	Work that is likely to have an impact on safety, environment, quality, or shut down the operation
3	Normal	Work should start within 48 h	Work that is likely to impact the production within a week.
4	Scheduled	As scheduled	Preventive maintenance and routine. All programmed work
5	Postponable	Work should start when resources are available or at shutdown period	Work that does not have an immediate impact on safety, health, environment, or the production operations

11.6 Scheduling Techniques

Scheduling is one of the areas that received considerable attention from researchers as well as practitioners in all types of applications including operations scheduling and project scheduling. Techniques are developed to develop optimum or near optimal schedules with respect to different possible performance measures. This chapter highlights some of these techniques and their application in maintenance scheduling.

11.6.1 Gantt Charts and Scheduling Theory

One of the oldest techniques available for sequencing and scheduling operations is the Gantt chart developed by Henry L. Gantt during World War II. The Gantt chart is a bar chart that specifies the start and finish time for each activity on a horizontal time scale. It is very useful for showing planned work activities vs accomplishments on the same time scale. It can also be used to show the interdependencies among jobs, and the critical jobs that need special attention and effective monitoring. There are large variations of the Gantt chart. To demonstrate the use of the Gantt chart several examples are given below. The example in Figure

11.5 shows the simplest form of the Gantt chart in which activities are scheduled at specified dates within the month.

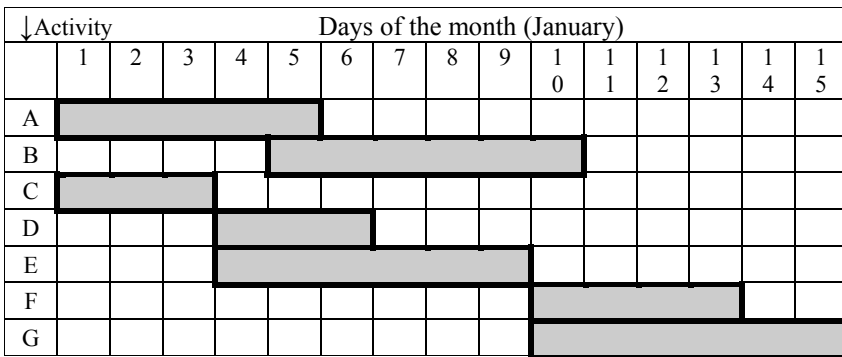
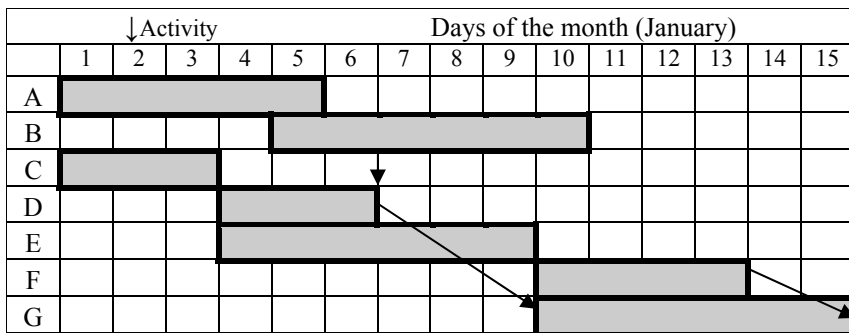


Figure 11.5. A Gantt chart representing a schedule of seven activities



Source: Duffuaa et al. (1999)

Figure 11.6. A Gantt chart with milestones

The example in Figure 11.5 modified to show interdependencies by noting milestones on each job timeline is shown in Figure 11.6. The milestones indicate key time periods in the duration of each job. Solid lines connect interrelationships among milestones. The milestones thus indicate the interdependencies between jobs. Obvious milestones for any job are the starting time for the job and the required completion point. Other important milestones are significant points within a job, such as the point at which the start of other jobs is possible.

Gantt charts can also be used to show the schedule for multiple teams or equipment simultaneously. A case in which three heavy pieces of equipment are scheduled for different jobs throughout the day is shown in Figure 11.7. The actual progression indicated in the chart shows any deviation from the scheduled timing. The chart indicates that jobs 25A and 15D are completed on schedule, job 25C is behind schedule by about a full day while job 25B is ahead of schedule by about a

day, and job 41E is in progress exactly on schedule. Jobs 33C and 44E scheduled but have not started yet.

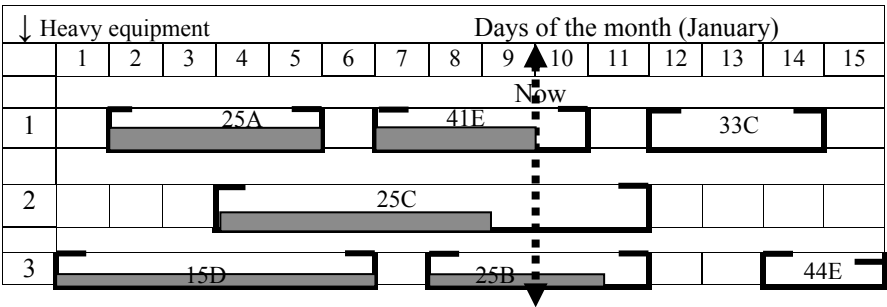


Figure 11.7. Gantt chart with progression

Color codes are sometimes used to reflect certain conditions such as shortage of material or machine breakdowns. Several scheduling packages, such as Primavera, are available to construct Gantt charts for more complicated schedules involving multiple resources and large number of activities. In general, Gantt chart does not build a schedule but helps in presenting the schedule in a simple visible manner that might help in monitoring, controlling and may be adjusting schedules. Scheduling (adding new jobs to the Gantt chart) itself is done following a certain rule that is developed with experience for the schedule to perform in the desired way. An example of such a rule is loading the heaviest job to the least loaded equipment as early as possible for maximizing the utilization of the equipment. This rule is known from scheduling theory to produce a good schedule for minimizing idle time.

Optimization techniques are available in the literature for such cases and for other cases with multiple or single resource. In general, scheduling theory has developed to handle short term production scheduling in different shop structures including job shop, flow shop, open shop and parallel machine structures See Pinedo (2002) for one of the recent books in scheduling theory. Integer programming is commonly used for developing optimum schedules for various scheduling requirements under various problem structures. However, they turn out to be large scale models that are quite complicated for real life situations. Another line of research in scheduling theory is developing heuristic methods, some of which are quite simple and practical, that result in good schedules with respect to certain performance measures. Computer simulation is heavily used in testing the performance of different competing heuristics and dispatching rules under stochastic system behavior including machine breakdowns, and stochastically dynamic job arrivals.

Some of the simple rules that can be utilized in maintenance scheduling are:

- For minimizing the average job waiting time, select jobs with high priority and short time requirements to be scheduled first. More specifically jobs

should be ordered in increasing order of the ratio of processing time to weighted job priority (assuming high priority jobs have high weights). This rule is known as the weighted shortest processing time (WSPT) rule in scheduling theory.

- For minimizing the average job waiting time having more than one team (crew) of the same capabilities, construct the schedule by assigning the job with the least time requirement to the fastest team.
- Having teams of different capabilities serving for different tasks for interrelated jobs (job shop environment), each team should select the task belonging to the job with the most remaining time requirement. This will maximize the utilization of maintenance crew (or equipment).

In spite of the developments in scheduling theory, its use in maintenance scheduling is limited due to the different nature of maintenance activities compared to production activities in many aspects including:

- Maintenance activities are highly uncertain in terms of duration and resource requirements;
- Maintenance activities are highly related in terms of precedence relations or relative priority;
- Tasks can be divided into subtasks each with different requirements; and
- Tasks can be interrupted or canceled due to changes in production conditions or maintenance requirements.

Recent advances in scheduling theory tended to tackle problems that are more stochastic in nature and some research is devoted to maintenance scheduling applications. Another recent trend in scheduling theory is the integration of maintenance scheduling and production scheduling which are traditionally done independently.

11.6.2 Project Scheduling

Maintenance activities commonly take the form of a project with many dependent operations forming a network of connected operations. In such cases, project management techniques can be utilized for scheduling the maintenance operations. The two primary network programming techniques used in project scheduling are the critical path method (CPM) and program evaluation and review technique (PERT). Each was developed independently during the late 1950s. The main difference between the two is that CPM uses a single estimate of activity time duration while PERT uses three estimates of time for each activity. Hence, CPM is considered to be a deterministic network method while PERT is a probabilistic method. Both networks consist of nodes representing activities and arrows indicating precedence between the activities. Alternatively, arrows may represent activities and nodes represent milestone. Both conventions are used in practice. Here we are going to use the former.

The objective in both CPM and PERT is to schedule the sequence of work activities in the project and determine the total time needed to complete the project. The total time duration is the longest sequence of activities in the network (the longest path through the network diagram) and is called the critical path. Before we

proceed by explaining the two methods it is worth noting that PERT and CPM are not well suited for day-to-day independent small jobs scheduling in a maintenance department. However, they are very useful in planning and scheduling large jobs (20 man hours or more) that consist of many activities such as machine overhauls, plant shut downs, and turnaround maintenance activities. Furthermore, a prerequisite for the application of both methods is the representation of the project as a network diagram, which shows the interdependencies and precedence relationships among the activities of the project.

Formulating the maintenance project as a network diagram helps in viewing the whole project as an integrated system. Interaction and precedence relationships can be seen easily and be evaluated in terms of their impact on other jobs. The project network representation will be demonstrated by an example from maintenance. Table 11.2 shows the data for overhauling a bearing in a train cargo carriage. The data shows the normal, crash duration, their corresponding costs, and precedence relationships for each activity. The term crash time refers to the minimum time the job can be accomplished in (by committing more resources), beyond which no further reduction in the job duration can be achieved. At this duration any increase in the resources for this job will increase the cost without reducing the duration.

Table 11.2. Normal and crash data for bearing overhaul

Activity	Description	Time (Min.)		Costs(\$)		Immediate precedence relationship
		Normal	Crash	Normal	Crash	
A	Dismantling	50	30	100	150	0
B	Repair of bolster pockets	67	50	120	150	A
C	Repair side frame rotation stop legs	90	60	150	200	A
D	Check friction blocks and all springs	35	25	50	75	A
E	Repair bolster rotation stop gibs.	35	25	140	170	B
F	Repair side frame column wear plates	55	40	100	130	C
G	Repair bolster pivot	210	150	250	300	E
H	Assemble	65	45	120	150	D, F and G
I	Painting	40	30	80	100	H

Source: Duffuaa et al. (1999)

Figure 11.8 shows the network corresponding to the data in the table. It starts with node A with no predecessor activity and it is represented by a circle nearby a number indicating the time. A itself is a predecessor for three activities B, C, and D drawn as three circles connected to A by arrows to indicate the precedence relation with A. Other activities (nodes) are traced back similarly. The resulting network is terminated by node I that has no successor.

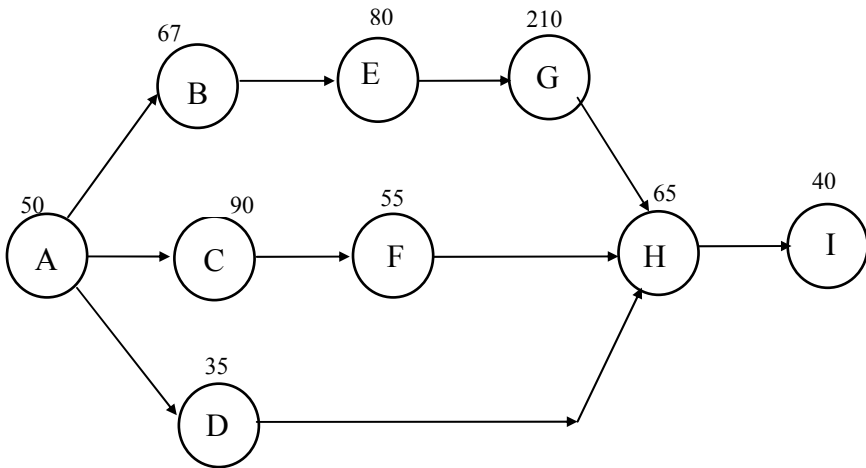


Figure 11.8. Network diagram for bearing overhaul data

There are many paths through the network in Figure 11.8 starting from the first node to the last node. The longest one is called the critical path and the summation of the activity times along that path is the total project duration. Jobs in the critical path are called critical in the sense that any delay in these jobs would cause a delay in the whole project. All other paths include slack times (sometimes called floats), *i.e.*, the amount of extra time that activities in the path can be delayed without delaying the completion time of the whole project. Activities that are not in the critical path may have some slack times, *i.e.*, delaying this activity for one reason or another will not delay the whole project. In this example there are three possible paths shown in Table 11.3. Critical activities must be monitored carefully and adhere to their specified schedules; however, non-critical activities can be used for leveling the resources due to the available slacks.

Table 11.3. Possible paths for completing bearing overhaul

Path	Path activities	Project duration	Sum
1	A-B-E-G-H-I	$50+67+35+210+65+40$	467
2	A-C-F-H-I	$50+90+55+65+40$	300
3	A-D-H-I	$50+35+65+40$	190

Clearly the project duration is 467 min and the critical path is the first path (A-B-E-G-H-I). Paths 2 and 3 have slacks of 167 and 277 min respectively. In this example, it was easy to go through all possible paths to find the one with the longest time; however, it would be extremely difficult to do the same for larger projects having a large number of activities and more complicated relationships

between them. A systematic approach for identifying the critical path is known as the critical path method (CPM).

11.6.3 Critical Path Method

To identify the critical path using the CPM method we need to follow the following steps:

- 1. Develop the project network diagram as shown in the previous section;
- 2. Perform the CPM calculation to identify the critical jobs (there are jobs on the critical paths and non-critical jobs (which are jobs with float);
- 3. Perform project crashing to (determine minimum times for each job) reduce project duration and investigate the cost tradeoffs; and
- 4. Level the resources in order to have uniform manpower requirements to minimize hiring, firing, or overtime requirements.

The critical path calculation includes two phases. The first phase is the forward pass (starting with the first node and proceeding to the last node). In this phase, the earliest start time, ES, and earliest finish time, EF, are determined for each activity. The earliest start time ES_i for a given activity, i , is the earliest possible time in the schedule that activity i can be started. Its value is determined by summing up the activity times of the activities lying on the longest path leading to it. The earliest finish time EF_i for a given activity i , is its earliest start time plus its activity time T_{ai} . The calculations for the bearing overhaul example are shown in Table 11.4.

Table 11.4. Earliest start times and finish times for the example

Activity I	Longest forward path	ES_i	T_{ai}	EF_i
A	-	0	50	50
B	A	50	67	117
C	A	50	90	140
D	A	50	35	85
E	A-B	117	35	152
F	A-C	140	55	195
G	A-B-E	152	210	262
H	A-B-E-G	362	65	427
I	A-B-E-G-H	427	40	467

The second phase is the backward pass (starting with the last node and proceeding back to the first node). We start this phase by assuming that the total project time T_{cp} , is the earliest finish time, EF, of the last activity found in the forward pass. In this phase, the latest finish time, LF, and latest start time, LS, are determined for each activity. The latest finish time LF_i for a given activity, i , is the latest possible time that activity i must be completed in order to finish the whole project on schedule. Its value is determined by subtracting from T_{cp} the activity

time along the longest path leading backward from the last node. For the last activity of the schedule, LF is set to be the total time duration of the project, T_{cp} . The latest finish time, LF_i , for a given activity, i , is its latest finish time minus its activity time T_{ai} . The calculations for the bearing overhaul example are shown in Table 11.5.

Table 11.5. Latest finish times and start times for the example

Activity i	Longest forward path	Length of the longest path	LF_i ($T_{cp}=467$)	T_{ai}	LS_i
I	-	0	467	40	427
H	I	40	427	65	362
G	I-H	105	362	210	152
F	I-H	105	362	55	307
E	I-H-G	315	152	35	117
D	I-H	105	362	35	327
C	I-H-F	160	362	90	372
B	I-H-G-E	350	117	67	50
A	I-H-G-E-B	417	50	50	0

The last step in the analysis of the network is to determine the slack time for each activity S_i . It can be determined by the difference between the latest and the earliest start time of the activity. The calculations are shown in Table 11.6 below.

Table 11.6. Slack times for the example

Activity i	LS_i	ES_i	LF_i	EF_i	S_i
A	0	0	50	50	0
B	50	50	117	117	0
C	372	50	362	140	322
D	327	50	362	85	277
E	117	117	152	152	0
F	307	140	362	195	167
G	152	152	362	262	0
H	362	362	427	427	0
I	427	427	467	467	0

Note that the activities along the critical path (A-B-E-G-H-I) have zero slack times. Activities not lying on the critical path have positive slacks, meaning that they could be delayed by an amount of time equal to their slack without delaying the project completion time.

The construction of the time chart should be made taking into consideration the available resources, and must take full advantage of the CPM calculation. In some circumstances it might not be possible to schedule many activities simultaneously because of personnel and equipment limitations. The total float for non-critical activities can be used to level the resources and minimize the maximum resource requirement. These activities can be shifted backward and forward between maximum allowable limits and scheduled at an appropriate time that levels the resources and keeps a steady workforce and equipment.

In addition to resource leveling, CPM involves project crashing. In project crashing, the duration of one or more critical activities are shortened in an optimal fashion and a curve is prepared to show the trade off between time and cost. This will enable management to evaluate project duration with the resulting cost. Network programming can be used to perform crashing in an optimal fashion. For more on project scheduling, see Taha (1992).

11.6.4 Program Evaluation Review Techniques (PERT)

Maintenance activities are usually unique and commonly involve unexpected needs that make their time duration highly uncertain. CPM uses a single estimate of the time duration based on the judgment of a person. PERT, on the other hand, incorporates the uncertainty by three time estimates of the same activity to form a probabilistic description of their time requirement. Even though the three time estimates are judgmental they provide more information about the activity that can be used for probabilistic modeling. The three values are represented as follows:

O_i = optimistic time, which is the time required if execution goes extremely well;

P_i = pessimistic time, which is the time required under the worst conditions; and

m_i = most likely time, which is the time required under normal condition.

The activity duration is modeled using a beta distribution with mean (μ) and variance (σ^2) for each activity i estimated from the three points as follows:

$$\hat{\mu}_i = \frac{O_i + P_i + 4m_i}{6}$$

$$\hat{\sigma}_i^2 = \left(\frac{P_i - O_i}{6} \right)^2$$

Estimated means are then used to find the critical path in the same way of the CPM method. In PERT, the total time of the critical path is a random variable with a value that is unknown in advance. However, additional probabilistic analysis can be conducted regarding possible project durations based on the assumption that the total time of the project may be approximated by a normal probability distribution with mean μ and variance σ^2 estimated as

$$\hat{\mu} = \sum \hat{\mu}_i$$

$$\hat{\sigma}^2 = \sum \hat{\sigma}_i^2$$

where i is an activity in the critical path

Using the above approximation we can calculate the probability with which a project can be completed in any time duration, T , using the normal distribution as follows:

$$\Pr(T_{cp} \leq T) = \Pr(Z \leq \frac{T - \hat{\mu}}{\sqrt{\hat{\sigma}^2}}) = \Phi(z)$$

Where Φ is the distribution function of the standard normal distribution.

Tables exist for evaluating any probability under the standard normal distribution. To illustrate the PERT analysis, consider the previous example with additional time estimates shown in Table 11.7 below.

Table 11.7. The PERT calculation for the bearing overhaul example

Activity	Description	Time (min)			Estimates	
		O	m	P	Mean	Variance
A	Dismantling	40	50	60	50	11.09
B	Repair of bolster pockets	60	67	74	67	5.43
C	Repair side frame rotation stop legs	85	90	95	90	2.79
D	Check friction blocks and all springs	32	35	38	35	4.00
E	Repair bolster rotation stop gibs	30	35	40	35	2.79
F	Repair side frame column wear plates	50	55	60	55	2.79
G	Repair bolster pivot	170	210	250	210	177.69
H	Assemble	59	65	71	65	4.00
I	Painting	35	40	45	40	2.79
Total					467	213.37

The critical path calculations lead to the same critical path obtained in the previous CPM calculations. The total project time is expected to be 467 min. The estimated variance is 213.37 min. The probability that the project will complete in 467 min can be calculated from the standard normal distribution to be 0.5, or the project has a 50% chance of completing in 467 min. The probability that the project may finish in 500 min can be calculated as:

$$\Phi\left(\frac{500 - 467}{\sqrt{213.37}}\right) = \Phi(2.26) = 0.9881$$

meaning that, the chance of completing the project in 500 min is almost 99%.

11.7 Scheduling Using Computers

It is always desirable to have a scheduling system that matches required maintenance work to available personnel and necessary equipment. The system should help maintain information of all necessary data and make them available with high reliability to build working schedules that optimizes the utilization of human resources and heavy equipment. A large number of software packages are available for optimum scheduling of personnel for planned maintenance activities and that takes into account the possibility of unplanned maintenance activities. Project scheduling packages are available to perform various functions related to project management. One of the leading packages is Microsoft Project that has the capability of maintaining data and generating Gantt charts for the projects. The critical path through the network diagram is highlighted in color to allow schedule monitoring and test alternatives.

Enterprise Resource Planning (ERP) is increasingly adopted by large enterprises as a global information and data management system to integrate the information flow through various functions within, and sometimes, outside the enterprise. The maintenance function is highly influenced by other functions in the enterprise through information flow as well as strategic directions. ERP is therefore extremely useful for integrating maintenance with production, spare part inventory, and engineering and purchasing. For more details about maintenance strategy integration in ERP see Nikolopoulos *et al.* (2003).

11.8 Summary

Maintenance planning and scheduling must serve the global objectives in the enterprise; hence it must be based on clear vision of its role in its success. Maintenance strategic planning is the process that assures matching between the maintenance objectives and objectives of the whole enterprise as well the objectives of other functional objectives. It selects the appropriate strategies regarding service delivery mode and type of contracts for outsourcing if needed as well as the organization and work structure and maintenance management methodology. In view of the selected strategies, long, medium and short range plans are constructed for time spans ranging from one year in the long term to weekly plans in the short term. The plans are then translated to schedules for implementing the plans at all levels. Master schedules are developed for long range plans and short range schedules are developed for days or hours within a day. Techniques exist in the literature to assist the planner and the scheduler in constructing good plans and schedules that achieve the objectives in the most efficient way. Gantt charts are usually used to monitor and control schedules. Methods like CPM and PERT are used to schedule maintenance activities forming a single large size project.

References

- Bedworth DD, Bailey JE (1987) *Integrated production control systems*, 2nd edn. John Wiley NY
- Budai G, Dekker R, Nicolai RP, (2006) A review of planning models for maintenance and production. *Econometric Institute Report*.
<http://hdl.handle.net/1765/8022>.
- Campbell JD (1995) Outsourcing in maintenance management: a valid alternative to self provision. *J of Qual in Maint Eng* 1(3): 18–24.
- Campbell JD, Reyes-Picknell JV (2006) *Uptime: strategies for excellence in maintenance management*. 2nd edition, Productivity Press.
- Cassady RC, Kutanoglu E (2005) Integrating preventive maintenance planning and production scheduling for a single machine. *IEEE Trans on Reliab* 24(2): 304–309.
- Duffuaa SO, Al-Sultan KS (1997) Mathematical programming approaches for the management of maintenance engineering. *J of Qual in Maint Eng* 3(3): 163–176.
- Duffuaa SO, Raouf A, Campbell JD (1999) *Planning and control of maintenance systems: modeling and analysis*. Wiley, NY.
- Escudero LF (1982) On maintenance scheduling of production units. *Eur J of Oper Res* 9(3): 264–274.
- Garg A, Deshmukh SG (2006) Maintenance management: literature review and directions. *J of Qual in Maint Eng* 12(3): 205–238.
- Groover MP (2007) *Work systems and methods, measurement, and management of work*. Prentice Hall, NJ.
- Hartman EH (1986) Maintenance productivity: why it is so low and how to improve it. *National Productivity Review*, Summer 1986: 224–232.
- Hipkin IB, Cock CDE (2000) TQM and BPR: lessons for maintenance management. *Omega The Int J of Manag Sci* 28: 277–292.
- Martin HH (1997) Contracting out maintenance and a plan for future research. *J of Qual in Maint Eng* 3(2): 81–90.
- Moder JJ, Phillips CC, Davis EW (1983) *Project management with CPM, PERT, and precedence diagramming*. 3rd edn. Van Nostand Reinhold Company, NY.
- Montgomery DC (1976) *Forecasting and time-series analysis*. McGraw Hill, NY.
- Murthy DNP, Atrons A, Eccleston JA (2002) Strategic maintenance management. *J of Qual in Maint Eng* 8(4): 287–305.
- Niebel BW (1994) *Engineering maintenance management*. 2nd edn. Marcel Dekker, NY
- Nikolopoulos K, Metaxiotis K, Lekatis N, Assimakopoulos V (2003) Integrating industrial maintenance strategy into ERP. *Ind Manag and Data Syst* 103(3): 184–191.
- Palmer RD (1999) *Maintenance planning and scheduling handbook*. McGraw-Hill, NY.
- Paz NM, Leigh W (1994) Maintenance scheduling: issues, results and research needs. *Int J of Oper & Prod Manag* 15(8):47–69.
- Pinedo M (2002) *Scheduling theory, algorithms, and systems*. Prentice Hall, New Jersey.
- Roberts SM, Escudero LF (1983) Minimum problem-size formulation for the scheduling of plant maintenance personnel. *J of Optim Theor and Appl* 39(3): 345–362.
- Raza SA, Al-Turki UM (2007) A comparative study of heuristic algorithms to solve maintenance scheduling problem. *J of Qual in Maint Eng* 13(4): 398–410.
- Ruiz R, Garcia-Diaz C, Maroto C (2007) Considering scheduling and preventive maintenance in the flow shop sequencing problem. *Comput and Oper Res* 34(11): 3314–3330.
- Sherwin D (2000) A review of overall models for maintenance management. *J of Qual in Maint Eng* 6(3): 138–164.
- Taha H (1992) *Operations research: An introduction*. 5th edn. Maxwell Macmillan.

- Tsang AHC (1998) A Strategic approach to managing maintenance performance. *J of Qual in Maint Eng* 4(2): 87–94.
- Tsang AHC (2002) Strategic dimensions of maintenance management. *J of Qual in Maint Eng* 8(1): 7–39.
- Tu PYL, Yam R, Tse P, Sun AOW(2001) An integrated maintenance management system for advanced manufacturing company. *The Int J of Adv Manuf Technol* 17: 692–703.
- Turner WC, Mize JH, Case KE, Nazemetz JW (1993) *Introduction to industrial engineering*, 3rd edn. Prentice Hall, NJ.
- Visser JK (1998) Modelling maintenance performance: a practical approach. *IMA Conference*, Edinburgh, pp 1–13.
- Watson P (1998) Performance specified maintenance contracts – why it is better for a client to specify desired results rather than how to achieve them. *Proc. Of the 3rd Intl. Conf. of Maintenance Societies*, Adelaide, Australia, pp 1–9.
- Worrall BM, Mert B (1980) Application of dynamic scheduling rules in maintenance planning and scheduling. *Int J of Prod Res* 18(1): 57–71.

Models for Production and Maintenance Planning in Stochastic Manufacturing Systems

E.K. Boukas

12.1 Introduction

Production systems are the facilities by which we produce most of the goods we are consuming in our daily lives. These goods range from electronics parts to cars and aircrafts. The production systems are in general complex systems and represent a challenge for the researchers from operations research and control communities. Their modeling and control are among the hardest problems we can have.

In the literature, we can find two main approaches that have been used to tackle the control problems for manufacturing systems (see Aghezzaf Jamali and Ait-Kadi 2007; Boukas, 1998; Boukas and Haurie, 1990; Gershwin, 1993; Lejeune and Ruszczynski, 2007; Panogiotidou and Tagaras, 2007; Sethi and Zhang, 1994; Sharifnia *et al.* 1991; Yang *et al.* 2005) and references therein). The first one supposes that the production system is deterministic (neglecting all the random events that may occur) and uses either the linear programming or dynamic programming to solve the production planning problem (see Maimoun *et al.* 1998) and references therein). Some attempts to include the maintenance have also been proposed. The second approach includes the random events like breakdowns, repairs, *etc.* that are inevitable in such systems and uses either the control theory or operations research tools to deal with the production and the maintenance planning.

In the last decades the production and maintenance planning problem has been an active area of research. The contribution on this topic can be divided into two categories. The first one ignores the production planning and considers only the maintenance planning; for more details on this direction we refer the reader to Wang (2002) and references therein, while the second category combines the production and the maintenance planning (see Marquez *et al.* 2007) and references therein). For a recent review of maintenance policies that have been used for production systems we refer the reader to the recent survey on the topics by Wang (2002) and also to reference therein.

The aim of this chapter is to propose models that provide simultaneously the production and maintenance planning for manufacturing systems with random breakdowns. Two models are covered. The first uses the continuous-time framework and, based on the dynamic programming approach, the policies of production and maintenance are computed. The second uses discrete-time framework and proposes a hierarchical approach with two levels with an appropriate algorithm to compute the production and maintenance. At the two levels, the problems are formulated as linear programming problems.

The rest of this chapter is organized as follows. In Section 2, the production and maintenance planning is formulated. In Section 3, the approach that uses the dynamic programming is presented and the procedure to solve the problem is developed. In Section 4, the approach that uses the linear programming is presented and the hierarchical algorithm is developed to compute the production and maintenance policies.

12.2 Problem Statement and Preliminary Results

Let us consider a manufacturing systems with random breakdowns. The system is assumed to be composed of m unreliable machines producing p part types. Since the machines are unreliable, it results that the production capacity will change randomly which will make it difficult to respond in some cases to a given demand. The preventive maintenance is a way to keep the average system capacity in a desired range and therefore be able to respond to the desired demand. This requires good planning of the maintenance and at the same time the production.

The problem we will tackle in this chapter consists of determining the production and the maintenance policies we should adopt in order to satisfy the desired demand despite the random events that may disturb the production planning. This chapter will propose two ways to deal with the production and maintenance planning. The first approach that will be developed in Section 12.3, uses the continuous-time framework and, based on dynamic programming, proposes a way to compute the solution of the production and maintenance planning. This approach unfortunately needs a lot of numerical computations. To avoid this, another approach is proposed in Section 12.4 and uses a hierarchical algorithm with two levels. It separates the production and the maintenance at the two levels and treats them separately as linear programming optimization problems.

Before ending this section, let us recall some results that will be used in Section 12.3. Mainly, we recall the piecewise deterministic problem and its dynamic programming solution and the numerical method that can be used to solve the Hamilton Jacobi Bellman equation.

Let E be a countable set and ℓ be a function mapping E into $\mathbb{N}$, $\ell: E \rightarrow \mathbb{N}$. For each $\alpha \in E$, E_α^0 denotes a Borel set of $\mathbb{R}^{\ell(\alpha)}$, $E_\alpha^0 \subset \mathbb{R}^{\ell(\alpha)}$. Define

$$E^0 = \bigcup_{\alpha \in E} E_\alpha^0 = \{(\alpha, z) : \alpha \in E, z \in E_\alpha^0\},$$

which is a disjoint union of E_α^0 s. For each $\alpha \in E$, we assume the vector field

$$g^\alpha : E_\alpha^0 \rightarrow E_\alpha^0$$

is a locally Lipschitz continuous function, determining a flow $\phi_\alpha(x)$. For each $x = (\alpha, z) \in E^0$, define

$$t_*(x) = \begin{cases} \inf \{t > 0 : \phi_\alpha(t, z) \in \partial E_\alpha^0\}, \\ \infty & \text{if no such time exists,} \end{cases}$$

where ∂E_α^0 is the boundary of E_α^0 . Thus $t_*(x)$ is the boundary hitting time for the starting point x . If $t_\infty(x)$ denotes the explosion time of the trajectory $\phi_\alpha(\cdot, z)$, then we assume that $t_\infty(x) = \infty$ when $t_*(x) = \infty$, thus effectively ruling out explosions. Now define

$$\begin{aligned} \partial^\pm E_\alpha^0 &= \{z \in \partial E_\alpha^0 : z = \phi_\alpha(\pm t, \xi) \text{ for some } \xi \in E_\alpha^0, t > 0\}, \\ \partial_1 E_\alpha^0 &= \partial^- E_\alpha^0 \setminus \partial^+ E_\alpha^0, \\ E_\alpha &= E_\alpha^0 \bigcup \partial_1 E_\alpha^0. \end{aligned}$$

With these definitions, the state space and boundary of a piecewise deterministic Markov process (PDP) can be respectively defined as follows:

$$E = \bigcup_{\alpha \in E} E_\alpha, \quad \text{state space,} \quad (12.1)$$

$$\Gamma^* = \bigcup_{\alpha \in E} \partial^+ E_\alpha^0, \quad \text{boundary.} \quad (12.2)$$

Thus the boundary of the state space consists of all those points which can be hit by the state trajectory. The points on some ∂E_α^0 which cannot be hit by the state of the trajectory are also included in the state space. The boundary of E consists of all the active boundary points, points in ∂E_α^0 that can be hit by the state trajectory.

The evolution of a PDP taking values in E is characterized by its three local characteristics:

1. A Lipschitz continuous vector field $f^\alpha : E \rightarrow \mathbb{R}^n$, which determines a flow $\phi_\alpha(t, z)$ in E such that, for $t > 0$.

$$\frac{d}{dt} \phi_\alpha(t, z) = f^\alpha(t, z), \quad \phi_\alpha(0, z) = z, \quad \forall x = (\alpha, z) \in E.$$

2. A jump rate $q : E \rightarrow \mathbb{R}_+$, which satisfies that for each $x \in E$, there is an $E > 0$ such that

$$\int_0^E q(\alpha, \phi_\alpha(t, z)) dt < \infty.$$

3. A transition measure $Q : E \rightarrow \mathcal{P}(E)$, where $\mathcal{P}(E)$ denote the set of probability measures on E .

By using these characteristics, a right-continuous sample path $\{x_t : t > 0\}$ starting at $x = (\alpha, z) \in E$ can be constructed as follows: define

$$x_t = (\alpha, \phi_\alpha(t, z)), \text{ if } 0 \leq t < \tau_1,$$

where τ_1 is the realization of the first jump time T_1 with the following generalized negative exponential distribution:

$$P(T_1 > t) = \exp\left(-\int_0^t q(\alpha, \phi_\alpha(s, z)) ds\right).$$

Having realized $T_1 = \tau_1$, we have $x_{\tau_1^-} = (\alpha, \phi_\alpha(\tau_1, z))$ and the post-jump state x_{τ_1} which has the distribution given by

$$P((\alpha', z_{\tau_1}) \in A \mid T_1 = \tau_1) = Q(A, x_{\tau_1^-})$$

on a Borel set A in E .

Restarting the process at x_{τ_1} and proceeding recursively according to the same recipe, one obtains a sequence of jump-time realizations $\tau_1, \tau_2, \dots$. Between each two consecutive jumps, $\alpha(t)$ remains constant and $z(t)$ follows the integral curves of f^α . Considering this construction as generic yields the stochastic process $\{x_t : t \geq 0, x_0 = x\}$ and the sequence of its jump times $T_1, T_2, \dots$. It can be shown that x_t is a strong Markov process with right continuous, left-limited sample paths (see Davis, 1993).

Piecewise-deterministic processes include a variety of stochastic processes arising from engineering, operation research, management science, economics and inventory system *etc.* Examples are queuing systems, insurance analysis (see Dassios and Embrechts, 1989), capacity expansion (see Davis *et al.* 1987), permanent health insurance model (Davis, 1993), inventory control model (see Sethi and Zhang, 1994), production and maintenance model (see Boukas and Haurie, 1990). Due to its extensive applications, the optimal control problem has received considerable attention. Gatarek (1992), Costa and Davis (1989), and Davis (1993) have studied the impulse control of PDPs. In the context of nonsmooth analysis, Dempster (1991) developed the condition for the uniqueness of the solution to the associated HJB equation of PDPs optimal control involving Clarke generalized gradient. The existence of relaxed controls for PDPs was proved by Davis (1993). Soner (1986), and Lenhart and Liao (1988) used the viscosity solution to formulate the optimal control of PDPs. For more information on the optimal control of PDPs, the reader is referred to Davis (1993) and Boukas (1987).

In this chapter, the models for the production and maintenance control in manufacturing system that we are treating here can be presented as a special class of piecewise deterministic Markov processes without active boundary points in the state space and the state jump can be represented by a function g . The model can be described as follows:

$$\dot{z}(t) = f^{\alpha(t)}(z(t), u(t)), \quad \forall t \in [T_n, T_{n+1}), \quad (12.3)$$

$$z(T_n) = g^{\alpha(T_n)}(z(T_n^-)), \quad n = 0, 1, 2, \dots \quad (12.4)$$

where $z = [z_1, \dots, z_p]^T \in \mathbf{R}^p$, $u = [u_1, \dots, u_q]^T \in \mathbf{R}^q$ are respectively, the state and control vectors, $f^\beta = [f_1^\beta, \dots, f_p^\beta]^T$ and $g^\beta = [g_1^\beta, \dots, g_p^\beta]^T$ represent real valued vectors, and x^T denotes the transpose of x . The initial conditions for the state and for the jump disturbance, the mode, are respectively $z(0) = z^0 \in \mathbf{R}^p$ and $\alpha(0) = \beta_0 \in E$. The set E is referred as the index set.

$\alpha = \{\alpha(t) : t \geq 0\}$ represents a controlled Markov process with right continuous trajectories and taking values on the finite state space E . When the stochastic process $\alpha(t)$ jumps from mode β to mode β' , the derivatives in Equation 12.3 change from $f^\beta(z, u)$ to $f^{\beta'}(z, u)$. Between consecutive jump times the state of the process $\alpha(t)$ remains constant. The evolution of this process is completely defined by the jump rates $q(\beta, z, u)$ and the transition probabilities $\pi(\beta' | \beta, z, u)$. The set E is assumed to be finite. T_n (random variable) is the time of the occurrence of the n th jump of the process α . For each $\beta \in E$, let $q(\beta, z, u)$ be a bounded and continuously differentiable function. At the jump time T_n , the state z is reset at a value $z(T_n)$ defined by Equation 12.4 where $g^\beta(\cdot) : \mathbf{R}^p \mapsto \mathbf{R}^p$ is, for any value $\beta \in E$, a given function.

Remark 12.2.1. This description of the system dynamics generalizes the control framework studied in depth by Rishel (1975), Wonham (1971) and Sworder and Robinson (1974), etc. The generalization lies in the fact that the jump Markov disturbances are controlled, and also from the discontinuities in the z -trajectory generated by Equations 12.3-12.4.

For each $\beta \in E$, let $f^\beta(\cdot, \cdot) : \mathbf{R}^p \times \mathbf{R}^q \mapsto \mathbf{R}^p$ be a bounded and continuously differentiable function with bounded partial derivatives in z . Let $U(\beta)$, $\beta \in E$, (a closed subset of $\mathbf{R}^q$) denotes the control constraints. Any measurable function with values in $U(\beta)$, for each $\beta \in E$, is called an admissible control. Let $\mathcal{U}$ be a class of stationary control functions $u_\beta(z)$, with values in $U(\beta)$ defined on $E \times \mathbf{R}^p$, called the class of admissible policies. The continuous differentiability assumption is a severe restriction on the considered class of optimization problems, but it is the assumption which allows the simpler exposition that was given in

Boukas and Haurie (1988). Later, in the practical models, the restriction will be removed by introducing the notion of viscosity solution of the Hamilton-Jacobi-Bellman equation.

The optimal control problem may now be stated as follows: given the dynamical system described by Equations 12.3-12.4, find a control policy $u_\beta(z) \in U$ such that the expected value of the cost functional

$$J(\beta, z, u) = \mathbf{E}_u \left\{ \int_0^\infty e^{-\rho t} c(\alpha(t), z(t), u(t)) dt \mid \alpha(0) = \beta, z(0) = z \right\} \quad (12.5)$$

is minimized over U .

In Equation 12.5, ρ ($\rho > 0$) represents the continuous discount rate, and $c(\beta, \cdot, \cdot) : \mathbf{R}^p \times \mathbf{R}^q \mapsto \mathbf{R}^+$, $\beta \in E$, is the family of cost rate functions, satisfying the same assumptions as $f^\beta(\cdot, \cdot)$.

We now proceed to give more precise definition of the controlled stochastic process. Let (Ω, F) be a measure space. We consider a function $X(t, \omega)$ defined as

$$\begin{aligned} X : D \times \Omega &\mapsto E \times \mathbf{R}^p, D \subset \mathbf{R}^+, \\ X(t, \omega) &= (\alpha(t, \omega), z(t, \omega)) \end{aligned}$$

which is measurable with respect to $B_D \times F$ (B_D is a σ -field).

Let $F_t = \sigma\{X(s, \cdot) : s \leq t\}$ be the σ -field generated by the past observations of X up to time t . We now assume the following:

Assumption 12.2.1 The behavior of the dynamical system at Equation 12.3 and 12.4 under an admissible control policy $u_\beta(\cdot) \in U$ is completely described by a probability measure P_u on (Ω, F_∞) . Thus the process $X_u = (X(t, \cdot), F_t, P_u), t \in D$, is well defined. For a given $\omega \in \Omega$ with $z(0, \omega) = z^0$ and $\alpha(0, \omega) = \beta_0$, we define

$$\begin{aligned} T_1(\omega) &= \inf\{t > 0 : \alpha(t, \omega) \neq \beta_0\}, \\ \beta_1(\omega) &= \alpha(T_1(\omega), \omega), \\ &\vdots \\ T_{n+1}(\omega) &= \inf\{t > T_n(\omega) : \alpha(t, \omega) \neq \alpha(T_n, \omega)\}, \\ \beta_{n+1}(\omega) &= \alpha(T_{n+1}(\omega), \omega), \\ &\vdots \end{aligned}$$

Assumption 12.2.2 For any admissible control policy $u_\beta(\cdot) \in U$, and almost any $\omega \in \Omega$, there exists a finite number of jump times $T_n(\omega)$ on any bounded interval $[0, T]$, $T > 0$. Thus the function $X_u(t, \omega) = (\alpha_u(t, \omega), z_u(t, \omega))$ satisfies

$$\begin{aligned}
\alpha_u(0, \omega) &= \beta_0 \\
z_u(t, \omega) &= z^0 + \int_0^t f^{\beta_0}(z_u(s, \omega), u_{\beta_0}(z(s, \omega))) ds, \quad \forall t \in [0, T_1(\omega)), \\
&\vdots \\
\alpha_u(t, \omega) &= \beta_n(\omega) \\
z_u(t, \omega) &= g^{\beta_n(\omega)}(z_u(T_n^-(\omega), \omega)) + \int_{T_n(\omega)}^t f^{\beta_n(\omega)}(z_u(s, \omega), u_{\beta_n}(z(s, \omega))) ds, \\
&\quad \forall t \in [T_n(\omega), T_{n+1}(\omega)), \\
&\vdots
\end{aligned}$$

Assumption 12.2.3 For any admissible control policy $u_\beta(\cdot) \in \mathcal{U}$, we have:

$$\begin{aligned}
P_u(T_{n+1} \in [t, t+dt] \mid T_{n+1} \geq T_n, \alpha(t) = \beta_n, z(t) = z) &= q(\beta_n, z, u_{\beta_n}(z))dt + o(dt), \\
P_u(\alpha(t) = \beta_{n+1} \mid T_{n+1} = t, \alpha(t^-) = \beta_n, z(t^-) = z) &= \pi(\beta_{n+1} \mid \beta_n, x, u).
\end{aligned}$$

Given these assumptions and an initial state (β_0, z^0) , the question which will be addressed in the rest of this section is to find a policy $u_\beta(\cdot) \in \mathcal{U}$ that minimizes the cost functional defined by 12.5 subject to the dynamical system at Equation 12.3 and 12.4.

Remark 12.2.2. From the theory of the stochastic differential equations and the previous assumptions on the functions f^β and g^β for each β , we recall that the system at equation 12.3 and 12.4 admits a unique solution corresponding to each policy $u_\beta(z) \in \mathcal{U}$. Let $z^\beta(s; t, z)$ denote the value of this solution at time s .

The class of control policies $\mathcal{U}$ is such that for each β , the mapping $u_\beta(\cdot): z \rightarrow U(\beta)$ is sufficiently smooth. Thus for each control law $u(\cdot) \in \mathcal{U}$, there exists a probability measure P_u on $(\Omega, \mathcal{F})$ such that the process (α, z) is well defined and the cost (icost) is finite. Let the value function $V(\beta, z)$ be defined by the following equation:

$$V(\beta, z) = \inf_{u \in \mathcal{U}} \mathbf{E}_u \left\{ \int_0^\infty e^{-\rho\tau} c(\alpha(\tau), z(\tau), u(\tau)) d\tau \mid \alpha(0) = \beta, z(0) = z \right\}.$$

Under the appropriate assumptions, the optimality conditions of the infinite horizon problem are given by the following theorem:

Theorem 12.2.1 A necessary and sufficient condition for a control policy $u_\beta(\cdot) \in \mathcal{U}$ to be optimal is that for each $\beta \in \mathcal{E}$ its performance function $V(\beta, z)$ satisfies the nonlinear partial differential equation

$$\begin{aligned} \rho V(\beta, z) = \min_{u(\cdot) \in U(\beta)} \left\{ c(\beta, z, u) + \sum_{i=1}^p \frac{\partial}{\partial z_i} V(\beta, z) f_i^\beta(z(t), u_\beta(z)) - q(\beta, z, u_\beta(z)) V(\beta, z) \right. \\ \left. + \sum_{\beta' \in E - \{\beta\}} q(\beta, z, u) V(\beta' | \beta, z, u) \pi(\beta' | \beta, z, u) \right\}, \forall \beta \in E \end{aligned} \quad (12.6)$$

where $\frac{\partial}{\partial z_i} V(\beta, z)$ stands for the partial derivative of the value function $V(\beta, z)$ with respect to the component z_i of the state vector z .

Proof. The reader is referred to Boukas and Haurie (1988) for the proof of this theorem. $\square$

As we can see the system given by Equation 12.6 is not easy to solve since it combines a set of nonlinear partial derivatives equations and optimization problem. To overcome this difficulty, we can approximate the solution by using numerical methods. In the next section, we will develop two numerical methods to solve these optimality conditions and which we believe that they can be extended to other class of optimization problems especially the nonstationary case.

To approximate the solution of the Hamilton-Jacobi-Bellman (HJB) equation corresponding to the deterministic or the stochastic optimal control problem, many approaches have been proposed. For this purpose, we refer the reader to Boukas (1995) and Kushner and Dupuis (1992).

In this section we will give an extension of some numerical approximation techniques which were used respectively by Kushner (1977), Kushner and Dupuis (1992) and by Gonzales and Roffman (1985) to approximate the solution of the optimality conditions corresponding to other class of optimization problems. Kushner has used his approach to solve an elliptic and parabolic partial differential system associated with a stochastic control problem with diffusion disturbances. Gonzales and Roffman have used their approach to solve a deterministic control problem. Our aim is to use these approaches to solve a combined nonlinear set of coupled partial differential equations representing the optimality conditions of the optimization problem presented in last subsection. The idea behind these approaches consists, within a finite grid G_z^h with unit cell of lengths $(h_1, \dots, h_p)$ for the state vector and a finite grid G_u^h with unit cell of lengths $(y_1, \dots, y_q)$ for the control vector, of using an approximation scheme for the partial derivatives of the value function $V(\beta, z)$ which will transform the initial optimization problem to an auxiliary discounted Markov decision problem. This will allow us to use the well-known techniques used for this class of optimization problems such as successive approximation or policy iteration.

Before presenting the numerical methods, let us define the discounted Markov decision process (DMDP) optimization problem. Consider a Markov process X_t which is observed at time points $t = 0, 1, 2, \dots$ to be in one of possible states of some finite state space $\mathcal{S} = \{1, 2, \dots, N\}$. After observing the state of the process, an action must be chosen from a finite space action denoted by $\mathcal{A}$.

If the process X_t is in state s at time t and action a is chosen, then two things occur: (1) we incur a cost $c(s, a)$ which is bounded and (2) the next state of the system is chosen according to the transition probabilities $P_{ss'}(a)$.

The optimization problem assumes a discounted factor $\delta \in (0, 1)$, and attempts to minimize the expected discounted cost. The use of δ is necessary to make the costs incurred at future dates less important than the cost incurred today. A mapping $\gamma : \mathbf{S} \rightarrow \mathbf{A}$ is called a policy. Let $\mathbf{A}$ be set of all the policies. For a policy γ , let

$$V_\gamma(s) = \mathbf{E}_\gamma \left[\sum_{t=0}^{\infty} \delta^t c(X_t, a_t) \mid X_0 = s \right],$$

where $\mathbf{E}_\gamma$ stands for the conditional expectation given that the policy γ is used.

Let the optimal cost function be defined as

$$V_\alpha(s) = \inf_{\gamma} V(s).$$

In the following, we will recall some known results on this class of optimization problems. The reader is referred to Haurie and L'Ecuyer (1986) for more information on the topic and for the proofs of these results.

Lemma 12.2.1 The expected cost satisfies the following equation:

$$V_\alpha(s) = \min_{a \in \mathbf{A}} \left\{ c(s, a) + \delta \sum_{s'=1}^N P_{ss'}(a) V_\alpha(s') \right\}, \quad \forall s \in \mathbf{S}.$$

Let $B(I)$ denote the set of all bounded real-valued functions defined on the state space $\mathbf{S}$. Let the mapping T_α be defined by

$$\begin{aligned} T_\alpha : B(I) &\rightarrow B(I), \\ (T_\alpha w)(s) &= \min_{a \in \mathbf{A}} \left\{ c(s, a) + \delta \sum_{s'=1}^N P_{ss'}(a) w(s') \right\}, \quad \forall s \in \mathbf{S}. \end{aligned} \quad (12.7)$$

Let T_α^k be the composition of the map T_α with itself k times.

Lemma 12.2.2 The mapping T_α defined by Equation 12.7 is contractive.

Lemma 12.2.3 The expected cost $V_\alpha(\cdot)$ is the unique solution of the following equation:

$$V_\alpha(s) = \min_{a \in \mathbf{A}} \left\{ c(s, a) + \delta \sum_{s'=1}^N P_{ss'}(a) V_\alpha(s') \right\}, \quad \forall s \in \mathbf{S}.$$

Furthermore, for any $w \in B(I)$ the mapping $T_\alpha^n w$ converges to V_α as n goes to infinity.

Let us now see how we can put our optimization problem in this formalism. Since our problem has a continuous state vector z and a continuous control vector

u , we need first to choose an appropriate discretization of the state space and the control space. Let G_z^h and G_u^h denote respectively the corresponding discrete state space and discrete control space and assume that they have finite elements with respectively n_z points for G_z^h and n_u points for G_u^h .

For the mode of the piecewise deterministic system, we do not need any discretization. Let $\mathbf{S}$ denote the global state space, $\mathbf{S} = \mathbf{E} \times G_z^h$ and N its number of elements. As we will see later, the constructed approximating Markov process X_t will jump between these states, $(s = (\alpha, z) \in \mathbf{S})$, with the transition probabilities $P_{ss'}(a)$, when the control action a is chosen from G_u^h . These transition probabilities are defined as

$$P_{ss'}(a) = \begin{cases} p_h^\beta(z, z+h; a), & \text{if } z \text{ jumps} \\ \tilde{p}_h^\beta(\beta, z; \beta', a), & \text{if } \alpha \text{ jumps,} \end{cases}$$

where $p_h^\beta(z, z+h; a)$ and $\tilde{p}_h^\beta(\beta, z; \beta', a)$ are the probability transition between state s when the action a is used. The corresponding instantaneous cost function $c(s, a)$ and the discount factor δ of the approximating DMDP depend on the used discretization approach. Their explicit expressions will be defined later.

Let h_i denote the finite difference interval, in the coordinate i , and e_i the unit vector in the i th coordinate direction. The approximation that we use for $\frac{\partial}{\partial z_i} V(\beta, z)$ for each $\beta \in \mathbf{E}$, will depend on the sign of $f_i^\beta(z, u)$. Let G_z^h denote the finite difference grid which is a subset of $\mathbf{R}^p$.

This approach was used by Kushner to solve some optimization problems and it consists of approximating the value function $V(\beta, z)$ by a function $V_h(\beta, z)$, and to replace the first derivative partial derivative of the value function, $\frac{\partial}{\partial z_i} V(\beta, z)$, by the following expressions:

$$\frac{\partial}{\partial z_i} V(\beta, z) = \begin{cases} \frac{1}{h_i} [V_h(\beta, z + e_i h_i) - V_h(\beta, z)] & \text{if } \dot{z}(t) \geq 0 \\ \frac{1}{h_i} [V_h(\beta, z) - V_h(\beta, z - e_i h_i)] & \text{otherwise.} \end{cases} \quad (12.8)$$

For each β , define the functions $p_h^\beta(., ., .)$, $\tilde{p}_h^\beta(., ., .)$ and $Q_h^\beta(., .)$ respectively as follows:

$$Q_h^\beta(z, u) = q(\beta, z, u) + \sum_{i=1}^p [\dot{z}_i(t) \mid / h_i],$$

$$p_h^\beta(z; z \pm e_i h, u) = f_i^\pm(z, u) / [h_i Q_h^\beta(z, u)],$$

$$\tilde{p}_h^\beta(\beta, z; \beta', u) = q(\beta, z, u) \pi(\beta' \mid \beta, z, u) / Q_h^\beta(z, u),$$

$$f_i^+(z, u) = \max(0, f_i^\beta(z, u)),$$

$$f_i^-(z, u) = \max(0, -f_i^\beta(z, u)).$$

Let $p_h^\beta(z; z \pm h, u) = 0$ for all points z not in the grid.

Putting the finite difference approximation of the partial derivatives as defined in Equation 12.8 into Equation 12.6, and collecting coefficients of the terms $V_h(\beta, z)$, $V_h(\beta, z \pm e_i h)$, yields, for a finite difference interval h applying to z ,

$$V_h(\beta, z) = \left\{ \frac{c(\beta, z, u)}{Q_h^\beta(z, u) \left[1 + \frac{\rho}{Q_h^\beta(z, u)} \right]} + \frac{1}{\left[1 + \frac{\rho}{Q_h^\beta(z, u)} \right]} \left[\sum_{z' \in G_h} p_h^\beta(z; z', u) V_h(\beta, z') \right. \right. \\ \left. \left. + \sum_{\beta' \in \mathbf{E} - \{\beta\}} \tilde{p}_h^\beta(\beta, z; \beta', u) V_h(\beta' \beta''(z)) \right] \right\}. \quad (12.9)$$

Let us define $c(s, u)$ and δ as follows:

$$c(s, u) = \frac{c^\beta(z, u)}{Q_h^\beta(z, u) \left[1 + \frac{\rho}{Q_h^\beta(z, u)} \right]},$$

$$\delta = \frac{1}{1 + \frac{\rho}{Q_h^\beta(z, u)}}.$$

A careful examination of Equation 12.9 reveals that the coefficients of $V_h(\cdot, \cdot)$ are similar to transition probabilities between points of the finite set $\mathcal{S}$ since they are nonnegative and sum to, at most, unity. $c(s, u)$ is also nonnegative and bounded. δ , as defined, really represents a discount factor with values in $(0, 1)$. Then, Equation 12.9 has the basic form of the cost equation of the discounted Markov decision process optimization for a given control action. The approximating optimization problem built on the finite state space $\mathcal{S}$ has then the following cost equation:

$$V_h(\beta, z) = \min_{u \in G_u^h} \left\{ \frac{c(\beta, z, u)}{Q_h^\beta(z, u) \left[1 + \frac{\rho}{Q_h^\beta(z, u)} \right]} + \frac{1}{\left[1 + \frac{\rho}{Q_h^\beta(z, u)} \right]} \left[\sum_{z' \in G_h} p_h^\beta(z; z', u) V_h(\beta, z') \right. \right. \\ \left. \left. + \sum_{\beta' \in \mathbf{E} - \{\beta\}} \tilde{p}_h^\beta(z, \beta; \beta', u) V_h(\beta' \beta''(z)) \right] \right\}. \quad (12.10)$$

Based on the results presented previously, we claim the uniqueness and the existence of the solution of the approximating optimization problem. It is plausible that the algorithms used in the discounted Markov process optimization would be helpful in computing this solution.

12.3 Dynamic Programming Approach

Let us consider a manufacturing system that has m machines and produces n part types. When staying in stock, the produced parts of type j will deteriorate with constant rate $\gamma_j, 1 \leq j \leq n$. Suppose the machines are failure-prone and assume that every machine has p modes denoted by $S = \{1, \dots, p\}$. The mode of machine i is denoted by $r_i(t)$ and $r(t) = (r_1(t), \dots, r_m(t))^T \in \mathbf{S} = S^m$ denotes the state of the system. $r_i(t) = p$ means that machine i is under repair and $r_i(t) = j \neq p$ means that machine i is in mode j . In this mode, the machine can produce any part type with an upper production capacity $\bar{u}_j$. $r_i(t)$ is assumed to be a Markov process taking values in state space S with state transition probabilities

$$P[r_i(t+h) = l | r_i(t) = k] = \begin{cases} q_{kl}h + o(h), & \text{if } l \neq k \\ 1 + q_{kk}h + o(h), & \text{otherwise} \end{cases} \quad (12.11)$$

with $q_{kl} \geq 0$ for all $l \neq k$ and $q_{kk} = -\sum_{l \in S, l \neq k} q_{kl}$ for all $k \in S$, and $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$. Assume that $\{r_j(t), t \geq 0\}, 1 \leq j \leq m$ are independent. From these assumptions it follows that $\{r(t), t \geq 0\}$ is a Markov process, with state space $\mathbf{S}$ and generator $\Lambda = (\lambda_{\alpha\alpha'}), \alpha = (\alpha_1, \dots, \alpha_m), \alpha' = (\alpha'_1, \dots, \alpha'_m) \in \mathbf{S}$. These jump rates can be computed from the individual jump rates of the machines.

Suppose the demand rates of the products are constants and denoted by $d = (d_1, \dots, d_n)^T$. Let $u_{ij}(t)$ be the production rate of part type j on machine i and write

$$u(t) = \begin{pmatrix} u_{11}(t) & \cdots & u_{m1}(t) \\ u_{12}(t) & \cdots & u_{m2}(t) \\ \vdots & \vdots & \vdots \\ u_{1n}(t) & \cdots & u_{mn}(t) \end{pmatrix}$$

which are the control variables in this paper. To complete our model, let us give some notations. For any $x \in \mathbf{R}, x^+ = \max(x, 0), x^- = \max(-x, 0)$. For any $x \in \mathbf{R}^n$, let

$$x^\oplus = (x_1^+, \dots, x_n^+)^T, \quad x^\ominus = (x_1^-, \dots, x_n^-)^T, \quad |x| = (|x_1|, \dots, |x_n|)^T$$

and $\|x\|$ denote the Euclidian norm.

Under above assumptions, the differential equation that describes the evolution of the inventory of our facility is therefore given by

$$\dot{x}(t) = f(x(t), u(t), r(t)), x(0) = x_0, r(0) = \alpha, \quad (12.12)$$

where

$$f(x(t), u(t), r(t)) = -\gamma x^\oplus(t) + u(t)e - d, \quad (12.13)$$

with $\gamma = \text{diag}\{\gamma_1, \dots, \gamma_n\}$ and $e = (1, \dots, 1)^T \in \mathbb{R}^m$. In Equation 12.12 $u(t) \in \mathbb{R}_+^{n \times m}$ is the control vector which is assumed to satisfy the following constraints:

$$u(t) \in U(r(t)) = \Delta \left\{ u(t) : 0 \leq bu(t) \leq \bar{u}_{r(t)} \right\} \quad (12.14)$$

where $\bar{u}_{r(t)} = (\bar{u}_{r_1(t)}, \dots, \bar{u}_{r_m(t)})$ is the production capacity of the system and $b = (b_1, \dots, b_n)$ with $b_i \geq 0$ is a constant scalar.

Our objective is to seek a control law that minimizes the following cost function:

$$J(x_0, \alpha, u(\cdot)) = \mathbb{E} \left[\int_0^\infty e^{-\rho t} g(x(t), r(t)) dt \mid x(0) = x_0, r(0) = \alpha \right], \quad (12.15)$$

where ρ ($\rho \geq 0$) is the discount factor and $\mathbb{E}$ stands for the mathematical expectation operator, $g(x(t), r(t)) = [c^+ x^\oplus(t) + c^- x^\ominus(t)]$ with $c^+ \in \mathbb{R}_+^{1 \times n}$ being the inventory holding cost and $c^- \in \mathbb{R}_+^{1 \times n}$ is the shortage cost.

This optimization problem falls into the framework of the optimization of the class of systems with Markovian jumps. This class of systems has been studied by many authors and many contributions have been reported to the literature. Among them, we quote Krasovskii and Lidskii (1961), Rishel (1975), Boukas (1987), Sethi and Zhang (1994) and references therein.

The goal of the rest of this section is to determine what would be the optimal production rate $u(t)$ that minimizes the cost function at Equation 12.15. Before determining this control, let us introduce some useful definitions.

Definition 12.3.1 A control $u(\cdot) = \{u(t) : t \geq 0\}$ with $u(t) \in \mathbb{R}_+^{n \times m}$ is said to be *admissible* if (1) $u(\cdot)$ is adapted to the σ -algebra generated by the random process $r(\cdot)$, denoted as $\sigma\{r(s) : 0 \leq s \leq t\}$ and (2) $u(t) \in U(r(t))$ for all $t \geq 0$.

Let $\mathcal{U}$ denote the set of all admissible controls of our control problem.

Definition 12.3.2 A measurable function $u(x(t), r(t)) : \mathbb{R}^n \times \mathbb{S} \rightarrow \mathbb{R}_+^{n \times m}$ is an *admissible feedback control*, or simply the *feedback control*, if (1) for any given initial continuous state x and discrete mode α , the following equation has a unique solution $x(\cdot)$:

$$\dot{x}(t) = -\gamma x^\oplus(t) + u(x(t), r(t))e - d, \quad x(0) = x \quad (12.16)$$

and (2) $u(\cdot) = u(x(\cdot), r(\cdot)) \in \mathcal{U}$.

Let the value function $v(x(t), r(t))$ be defined by

$$v(x(t), r(t)) = \min_{u(t)} J(x(t), r(t), u(\cdot)). \quad (12.17)$$

Using the dynamic programming principle (see Boukas, 1987), we have

$$v(x(t), r(t)) = \min_{u(t)} \mathbb{E} \left[\int_t^\infty e^{-\rho(s-t)} g(x(s), r(s)) ds \mid x(t), r(t) \right]. \quad (12.18)$$

Formally, the Hamilton-Jacobi-Bellman equation can be given by the following:

$$\min_{u(t) \in \mathcal{U}(r(t))} [(A_u v)(x(t), r(t)) + g(x(t), r(t))] = 0, \quad (12.19)$$

where $(A_u v)(x(t), r(t))$ is defined as follows:

$$(A_u v)(x(t), r(t)) = f^{\gamma}(x(t), u(t), r(t)) \frac{\partial v}{\partial x}(x(t), r(t)) + \sum_{\beta \in \mathcal{S}} \lambda_{r(t)\beta} v(x(t), \beta). \quad (12.20)$$

To characterize the optimal control, let us establish some properties of the value function.

Theorem 12.3.1 For any control $u(\cdot) \in \mathcal{U}$, the state trajectory of Equation 12.12 has the following properties:

1. Let x_t be the state trajectory with initial state x_0 , then there exists

$$C_1 \in \mathbb{R}_+^n \text{ such that}$$

$$\|x_t\| \leq \|x_0\| + C_1 t. \quad (12.21)$$

2. Let x_t^1, x_t^2 be the state trajectories corresponding to $(x_1, u(\cdot))$ and $(x_2, u(\cdot))$ respectively, then there exists a constant $C_2 > 0$ such that

$$\|x_t^1 - x_t^2\| \leq C_2 \|x_1 - x_2\|, \quad (12.22)$$

implying

$$\|x_t^1 - x_t^2\| \leq C_2 \|x_1 - x_2\|.$$

Proof. For the proof of this theorem, we refer the reader to Boukas and Liu (2003).

Theorem 12.3.2 For each $r(t) \in \mathcal{S}$, the value function, $v(x(t), r(t))$, is convex;

1. There exists a constant C_3 , such that

$$v(x(t), r(t)) \leq C_3 (1 + \|x(t)\|);$$

2. For each $r(t) \in \mathcal{S}$, the value function, $v(x(t), r(t))$ is Lipschitz.

Proof. For the proof of this theorem, we refer the reader to Boukas and Liu (2003).

Theorem 12.3.3 Suppose that there is a continuously differentiable function $\hat{v}(x(t), r(t))$ which satisfies the Hamilton-Jacobi-Bellman equation at Equation 12.19. If there exists $u^*(\cdot) \in U$, for which the corresponding $x^*(t)$ satisfies at Equation 12.12 with $x^*(0) = x$, and

$$\min_{u \in U(r(t))} [(A_u \hat{v})(x^*(t), r(t))] = (A_{u^*} \hat{v})(x^*(t), r(t)) \quad (12.23)$$

almost everywhere in t with probability one, then $\hat{v}(x, \alpha)$ is the optimal value function and $u^*(\cdot)$ is optimal control,

$$\hat{v}(x, \alpha) = v(x, \alpha) = J(x, \alpha, u^*(\cdot)).$$

Proof. For the proof of this theorem we refer the reader to Boukas and Liu (2003).

This discussion shows that solving the optimal control problem involves solving HJB equation at Equation 12.19, which often doesn't have closed form solution in the general case. However, in the simplest case, Theorem 12.3.3 reveals that the optimal control has some special structure, which may be helpful to design the controller. In the sequel of this paper, we will restrict our study to the case of one machine that has two modes and produces one part type, $m=1, p=2, n=1, S=\{1, 2\}$. In this case, the deteriorating rate, production capacity and demand are denoted by $\gamma, \bar{u}$ and d respectively.

Let us also assume that the value function is continuously differentiable with respect to the continuous arguments. Using the expressions for the functions $f(\cdot)$ and $g(\cdot)$ and the HJB equation given by Equation 12.19, one has

$$\rho v(x, 1) = \min_u \left[(-\gamma x^+ + u - d) v_x(x, 1) + q_{11} v_x(x, 1) + q_{12} v_x(x, 2) + c^+ x^+ + c^- x^- \right] \quad (12.24)$$

$$\rho v(x, 2) = \min_u \left[(-\gamma x^+ - d) v_x(x, 2) + q_{21} v_x(x, 1) + q_{22} v_x(x, 2) + c^+ x^+ + c^- x^- \right] \quad (12.25)$$

Based on the structure of the optimality conditions, it results that the optimal control law is given by

$$u^*(t) = \begin{cases} \bar{u}, & \text{if } v_x(x(t), 1) < 0 \text{ and } r(t) = 1, \\ \gamma x^+(t) + d, & \text{if } v_x(x(t), 1) = 0 \text{ and } r(t) = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (12.26)$$

Moreover, by the convexity of $v(x, 1)$ we have

$$u^*(t) = \begin{cases} \bar{u}, & \text{if } x < x^*(t), \text{ and } r(t) = 1 \\ \gamma I_{\{x \geq 0\}} x^* + d, & \text{if } x = x^*(t) \text{ and } r(t) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (12.27)$$

where x^* is the minimal point of $v(x, 1)$, $v_x(x^*, 1) = 0$.

Let the optimal control be u^* and define

$$q_{12} = \lambda, \quad (12.28)$$

$$q_{21} = \mu, \quad (12.29)$$

$$V(x) = \begin{bmatrix} v(x, 1) \\ v(x, 2) \end{bmatrix}. \quad (12.30)$$

With these definitions and if we let x^* (without loss of generality, we assume x^* to be greater than 0, other cases can be handled similarly) denote the minimum of the value function at mode 1, the optimality conditions become

- $x > x^*$, then $u^* = 0$ and the optimality conditions become

$$V_x(x) = \begin{bmatrix} -\frac{\lambda+\rho}{\gamma x+d} & \frac{\lambda}{\gamma x+d} \\ \frac{\mu}{\gamma x+d} & -\frac{\mu+\rho}{\gamma x+d} \end{bmatrix} V(x) + \begin{bmatrix} \frac{c^+ x}{\gamma x+d} \\ \frac{c^+ x}{\gamma x+d} \end{bmatrix} \quad (12.31)$$

- $x = x^*$, then $\gamma x + d = u^*$ and the optimality conditions become

$$v(x, 1) = \left[\frac{\lambda}{\rho + \lambda} \right] v(x, 2) + \left[\frac{c^+ x}{\rho + \lambda} \right] \quad (12.32)$$

and

$$v_x(x, 2) = \left[-\frac{\rho(\rho + \lambda + \mu)}{(\rho + \lambda)(\gamma x + d)} \right] v(x, 2) + \left[\frac{\rho + \mu + \lambda}{\rho + \lambda} \right] c^+ x \quad (12.33)$$

- $0 \leq x < x^*$ then $u^* = \bar{u}$ and the optimality conditions become

$$V_x(x) = \begin{bmatrix} -\frac{\lambda+\rho}{\gamma x+d-\bar{u}} & \frac{\lambda}{\gamma x+d-\bar{u}} \\ \frac{\mu}{\gamma x+d} & -\frac{\mu+\rho}{\gamma x+d} \end{bmatrix} V(x) + \begin{bmatrix} \frac{c^+ x}{\gamma x+d-\bar{u}} \\ \frac{c^+ x}{\gamma x+d} \end{bmatrix} \quad (12.34)$$

- $x \leq 0$ then $u^* = \bar{u}$ and the optimality conditions become

$$V_x(x) = \begin{bmatrix} -\frac{\lambda+\rho}{d-\bar{u}} & \frac{\lambda}{d-\bar{u}} \\ \frac{\mu}{d} & -\frac{\mu+\rho}{d} \end{bmatrix} V(x) - \begin{bmatrix} \frac{c^-x}{d-\bar{u}} \\ \frac{c^-x}{d} \end{bmatrix} \quad (12.35)$$

To solve the HJB equations, we can use the numerical method used in Boukas (1995). This method consists of transforming the optimization problem to a Markov decision problem (MDP) with all the nice properties that guarantee the existence and the uniqueness of the solution. The key point of this technique is first to discretize the state space $\mathbf{R}$ and control space $[0, \bar{u}]$ to get a discrete state space $G_x = [-\underline{x}, -\underline{x} + h_x, \dots, \bar{x}]$ with $\underline{x}, \bar{x}$ great enough and a discrete control space $G_u = [0, h_u, \dots, \bar{u}]$, and then define a function $v_h(x, i)$ on $G_x \times S$ by letting $v_h(x, i) = v(x, i)$. By replacing $v_x(x, i)$ by

$$\begin{cases} \frac{1}{h_x} [v(x + h_x, i) - v(x, i)], & \text{if } f(x, u, i) \geq 0, \\ \frac{1}{h_x} [v(x + i) - v(x - h_x, i)], & \text{otherwise} \end{cases}$$

and substituting $v_h(x, i)$ into Equation 12.24 and 12.25 gives the following MDP problem:

$$v_h(x, 1) = \min_{u \in G_u} \left[c(x, 1) + \frac{1}{1 + \frac{\rho}{Q_{h_x}^1}} \left(\frac{(-\gamma x^+ + u - d)^+}{h_x Q_{h_x}^1} v_h(x + h_x, 1) + \frac{(-\gamma x^+ + u - d)^-}{h_x Q_{h_x}^1} v_h(x - h_x, 1) + \frac{q_{12}}{Q_{h_x}^1} v_h(x, 2) \right) \right], \quad (12.36)$$

$$v_h(x, 2) = c(x, 2) + \frac{1}{1 + \frac{\rho}{Q_{h_x}^2}} \left[\frac{|-\gamma x^+ - d|}{h_x Q_{h_x}^2} v_h(x - h_x, 2) + \frac{q_{21}}{Q_{h_x}^2} v_h(x, 1) \right], \quad (12.37)$$

where h_x is the discretization step for the x , $c(x, \alpha)$, $Q_{h_x}^1$ and $Q_{h_x}^2$ are defined by

$$c(x, \alpha) = \frac{c^+ x^+ + c^- x^-}{Q_{h_x}^\alpha \left[1 + \frac{\rho}{Q_{h_x}^\alpha} \right]}, \text{ for all } \alpha \in S, \quad (12.38)$$

$$Q_{h_x}^1 = \frac{|\gamma x^+ - u + d|}{h_x} + |q_{11}|, \quad (12.39)$$

$$Q_{h_x}^2 = \frac{|\gamma x^+ + d|}{h_x} + |q_{22}|. \quad (12.40)$$

The successive approximation technique and the policy iteration technique can be used to find an approximation of the optimal solution. For more information on

these techniques, we refer the reader to Bertsekas (1987), Boukas (1995) or Kushner and Dupuis (1992) and references therein.

Remark 12.3.1. By the same argument as in Boukas *et al.* (1996), it is easy to prove that $\lim_{h \rightarrow 0} v_h(x, i) = v(x, i), \forall i \in S$, which establishes the convergence of the approximation algorithm.

12.4 Linear Programming Approach

In the previous section we developed an approach to plan the production and maintenance using a continuous-time model. With this model we were able to compute simultaneously the production and maintenance. But this approach requires a lot of computations before the solution can be obtained. To overcome this, we propose a new approach that uses linear programming and an hierarchical algorithm for this purpose. To show how this approach works, we will restrict ourself to one machine one part type, but we have to keep in mind that the model we propose here is valid for any number of machines and part types. For this purpose, let us consider a production system with one machine that produces one part type and assume that the system must satisfy a given demand $d(k)$, $k = 0, 1, 2, \dots$ that can be constant or time varying. Let the dynamics of the production system be described by the following difference equation:

$$x(k) = x(k-1) + u(k) - d(k), x(0) = x_0 \quad (12.41)$$

where $x(k) \in \mathbb{R}$, $u(k) \in \mathbb{R}$ and $d(k) \in \mathbb{R}$ represent respectively the stock level, the production and the demand at period kT , $k = 0, \dots, N$.

The stock level, $x(k)$ and the production $u(k)$ must satisfy at each period kT the following constraints:

$$0 \leq u(k) \leq \bar{u} \quad (12.42)$$

$$x(k) \geq 0 \quad (12.43)$$

where $\bar{u}$ is known positive constant that represents the maximum production the system can have.

Remark 12.4.1. The upper bound constraint on the production represents the limitation of the capacity of the manufacturing system, while the one of the stock level means that we do not tolerate the negative stock. Notice that we can also include an upper bound of the stock level.

The objective is to plan the production in order to satisfy the given demand during a finite horizon. Since the capacity may change with time in a random way, it is required to include the preventive maintenance and combine it to the production planning problem. By performing maintenance we keep the capacity on average within certain acceptable values.

To solve the simultaneous production and maintenance planning problem, we use the following hierarchical approach with two levels:

1. At level one we plan the preventive maintenance; and
2. At level two, using the results of level one, we try satisfy the demand during the periods the machine is up.

To present each level in this algorithm let:

- T be the time period that can be 1 h, one day, 1 month, *etc.*;
- $x(k)$ be the stock level at time kT ;
- $u(k)$ be production at time kT ;
- $d(k)$ be the demand at time kT ;
- T_{up} be the amount of units of time during which the machine is working before the i th maintenance takes place (T_{up} is a multiple of T);
- T_d be the amount of units of time of the i th maintenance takes (T_d is a multiple of T and it is assumed to be the same for all the interventions);
- NT be the total time for the planning (N is a positive integer);
- ν be the upper bound of T_{up} ;
- μ be the number of preventive-maintenance taking place in NT ;
- $w_i(k)$ be the number of deferred items at time kT for i period;
- av be the availability of the machine;
- $\bar{u}$ be the upper bound of $u(k)$.

The algorithm we will adopt is summarized as follows:

1. Initialization: choose the data N, T, μ, T_d .
2. Solve a LP problem that gives the dates of the preventive intervention during the interval of time $[0, NT]$.
3. Test: if the problem is feasible go to Step 4, otherwise increase μ and go to Step 2.
4. Solve the LP problem for production planning to determine the decision variables.
5. Test: if the problem is feasible stop otherwise the interval of time $[0, NT]$ is not enough to respond to the demand and no feasible solution can be obtained. We can increase the interval and repeat the steps.

The problem at level one tries to divide the planning interval $[0, NT]$ in successive periods for production, T_{upk} and maintenance T_{dk} (T_{dk} is supposed to be constant here), $k = 1, 2, \dots, \mu$, that sum to a time that is less or equal to NT . It is also considered that the availability of the machine should be greater or to a given av . We should also note that T_{upk} is between 0 and ν for any k . The formulation of the optimization problem at this level is given by

$$\begin{cases} \min \max \left(\sum_{k=1}^{\mu} [T_{up_k} + T_d] \right) \\ s.t. : \\ \sum_{k=1}^{\mu} \{T_{up_k} + T_d\} \leq NT \\ \frac{\sum_{k=1}^{\mu} T_{up_i}}{\sum_{k=1}^{\mu} T_{up_i} + \mu T_d} \geq av \\ 0 \leq T_{up_i} \leq v \end{cases} \quad (12.44)$$

That can be transformed to

$$P1 : \begin{cases} \min Z \\ s.t. : \\ \sum_{k=1}^{\mu} T_{up_i} \leq Z - \mu T_d \\ \sum_{k=1}^{\mu} [T_{up_k} + T_d] \leq N \\ \frac{\sum_{k=1}^{\mu} T_{up_i}}{\sum_{k=1}^{\mu} T_{up_i} + \mu T_d} \geq av \\ 0 \leq T_{up_i} \leq v \end{cases} \quad (12.45)$$

which is a linear programming problem that can be easily solved using the powerful existing tools for this purpose.

The optimization problem at level two consists of performing the production planning within the time during which the machine is up in order to satisfy the demand and all the system constraints by penalizing the stock level and the production with appropriate unit costs. This problem is given by

$$P2 : \begin{cases} \min \sum_{k=1}^N [c^x x(k) + c^u u(k)] \\ s.t. : \\ x(k) = x(k-1) + u(k) - d(k), x(0) = x_0 \\ u(k) \leq \bar{u} \\ u(k) \geq 0 \\ x(k) \geq 0 \end{cases} \quad (12.46)$$

which is also a linear programming optimization problem.

Both the problems at the two levels are linear which make them easier to solve with the existing tools and for high dimensions problems. This can include production systems with multiple machines multiple part types.

To show the validity of the approach of this section, let us consider the system with the data of Table 12.1.

Table 12.1. System data

T	N	T_d	v	μ	c^x	c^u	$\bar{u}$
1	20	1	τ	4	2	3	4

Solving the previous optimization problems following the proposed algorithm with these data, we get the results of Figures 12.1–12.5.

Figure 12.1 of the machine gives the solution of the optimization problem at level one and it illustrates the sequence of the phases up and down for the considered machine. Since we don't impose conditions on the state of the machine when the age grows, the results at level one shows that we can perform periodic preventive maintenance that will take constant time as we did in this example.

Figure 12.2 shows the results of the solution of the optimization problem at level two for a given deterministic demand. This figure shows that the cumulative stock levels tracks well the given time-varying demand.

Figure 12.3 illustrated the production at each period obtained by the solution of the optimization problem at level two. As can be seen from these figures, all the constraints are satisfied.

With the same data, we have generated randomly the time-varying demand and solved the two levels optimization problems and the solution is illustrated by Figures 12.4–12.5.

In some circumstances due to reduction in the system capacity, we may defer the demand by some periods and pay a penalty cost. As first extension of the previous model let us now add the ability of deferring some items in the demand to the next period and see how to solve the production and maintenance planning for this case. First notice that the optimization problem at level one will not change since it is independent of the demand. The changes will mainly affect the second optimization problem, more specifically the cost function should take care of the cost incurred by the deferred items and the dynamics that must be changed to include the deferred items. The rest of the constraints on the stock level and the production stay unchanged. The changes in this case are:

- The previous cost function can be changed by the following: $\sum_{k=1}^N [c^x x(k) + c^u u(k) + c^w w(k)]$; and
- The new dynamics is

$$x(k) = x(k-1) + u(k) + w(k) - w(k-1) - d(k)$$

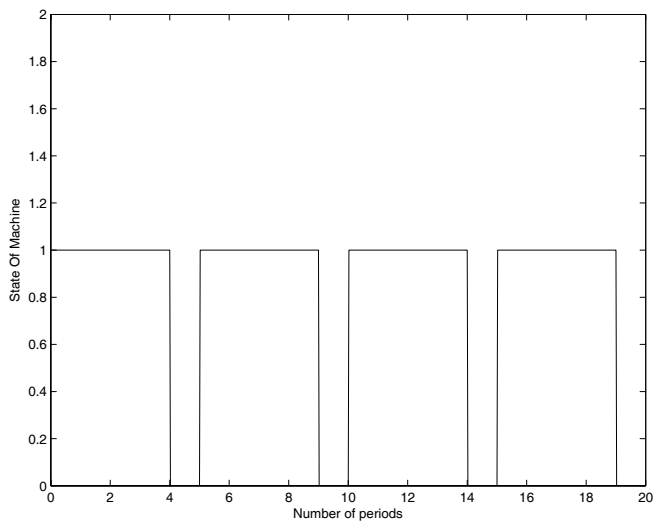


Figure 12.1. State of the machine

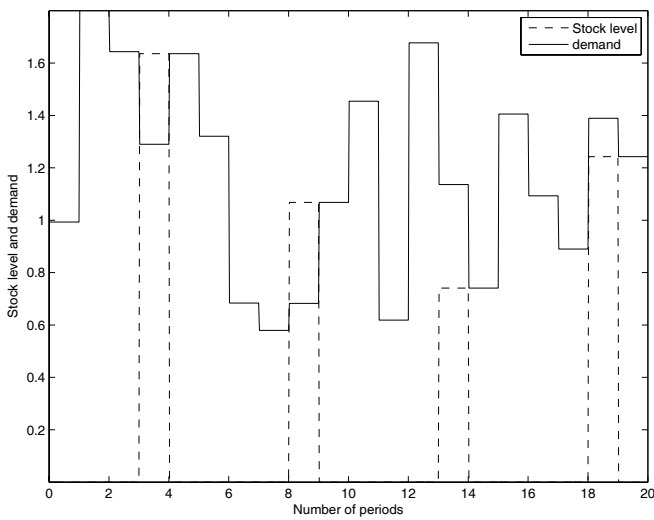


Figure 12.2. Stock level and demand (deterministic case)

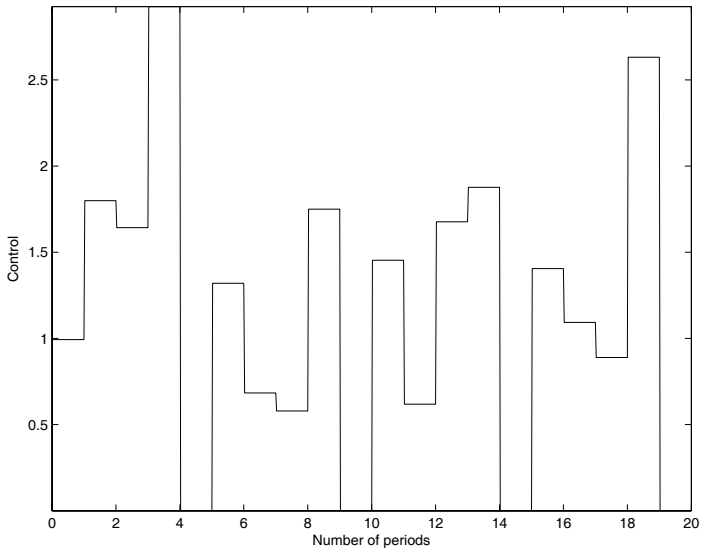


Figure 12.3. Production rate (deterministic case)

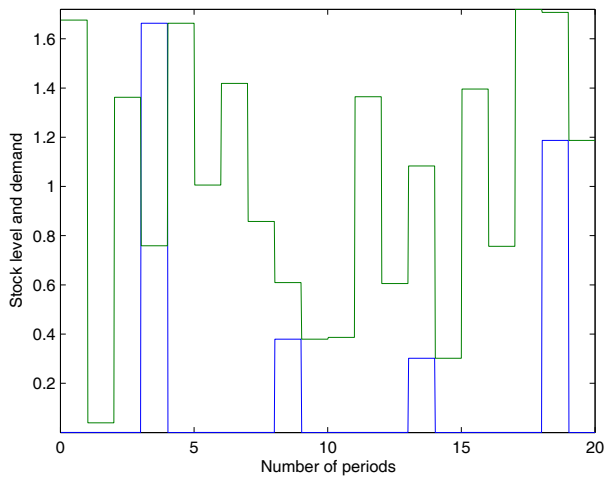


Figure 12.4. Stock level and demand (stochastic case)

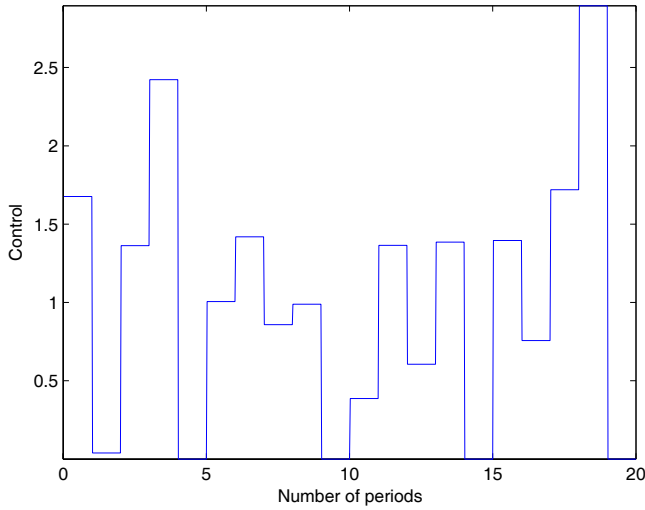


Figure 12.5. Production rate (stochastic case)

The optimization problem at level two becomes

$$P2' \begin{cases} \min \sum_{k=1}^N [c^x x(k) + c^u u(k) + c^w w(k)] \\ s.t. : \\ x(k) = x(k-1) + u(k) + w(k) - w(k-1) - d(k), x(0) = x_0 \\ u(k) \leq \bar{u} \\ u(k) \geq 0 \\ x(k) \geq 0 \end{cases} \quad (12.47)$$

With the same data of Table 12.1 solving the optimization problems at the two level, we get the results illustrated by Figures 12.6–12.11. Figure 12.6 gives the same results as for the case without deferred items. The other figures give the stock level and the production at different periods for the deterministic case and the stochastic one as we did for the previous model.

As a second extension, let us now add the ability to defer some items of the demand up to three periods. For this case the changes we have to make to our second optimization problem at level two concern the cost and dynamics. These changes are:

- The cost function becomes

$$\sum_{k=1}^N [c^x x(k) + c^u u(k) + c^{w_1} w_1(k) + c^{w_2} w_2(k) + c^{w_3} w_3(k)]$$

- The dynamics become

$$\begin{aligned}
 x(k) = & x(k-1) + u(k) + w_1(k) - w_1(k-1) + w_2(k) - w_2(k-2) \\
 & + w_3(k) - w_3(k-3) - d(k)
 \end{aligned}$$

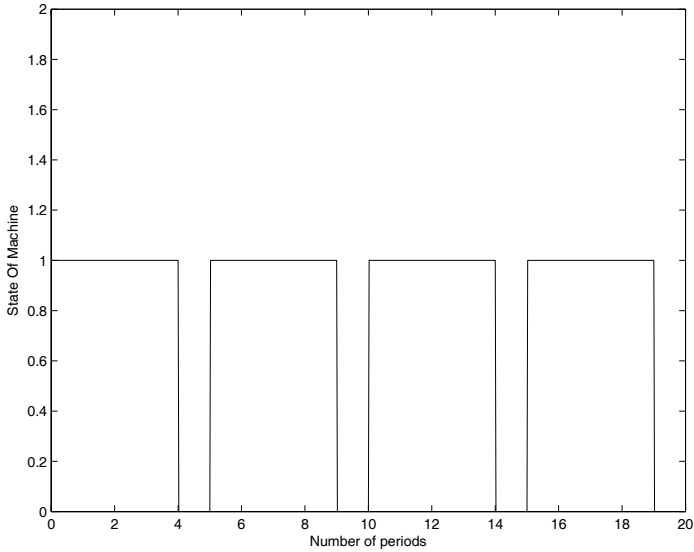


Figure 12.6. Stock level and demand (deterministic case)

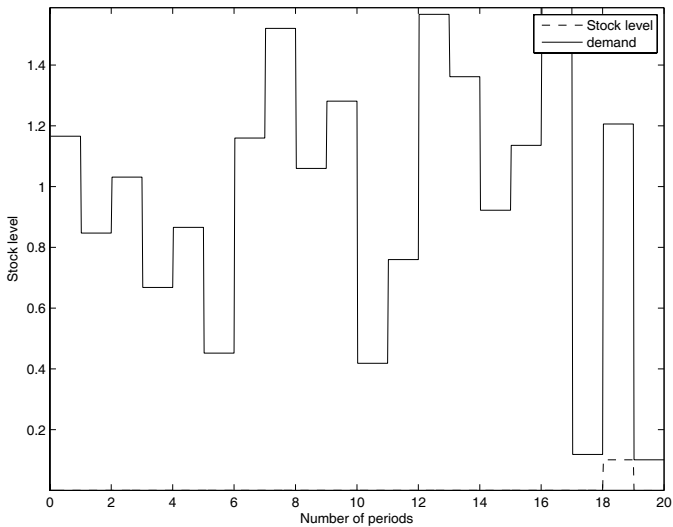


Figure 12.7. Production rate (deterministic case)

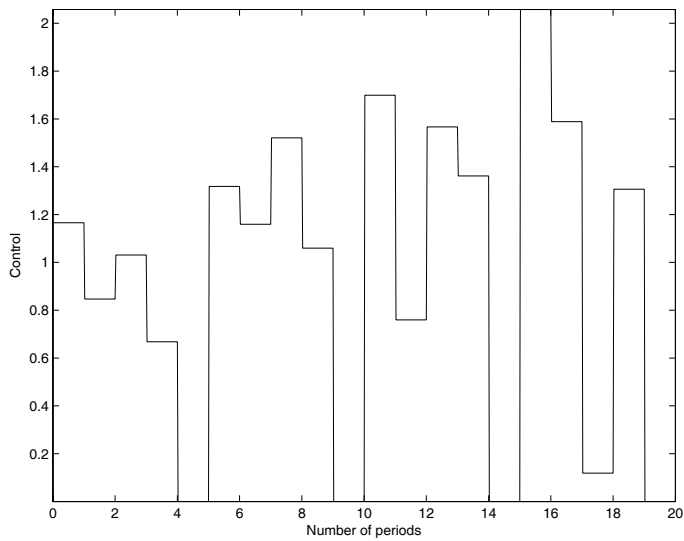


Figure 12.8. Deferred items (deterministic case)

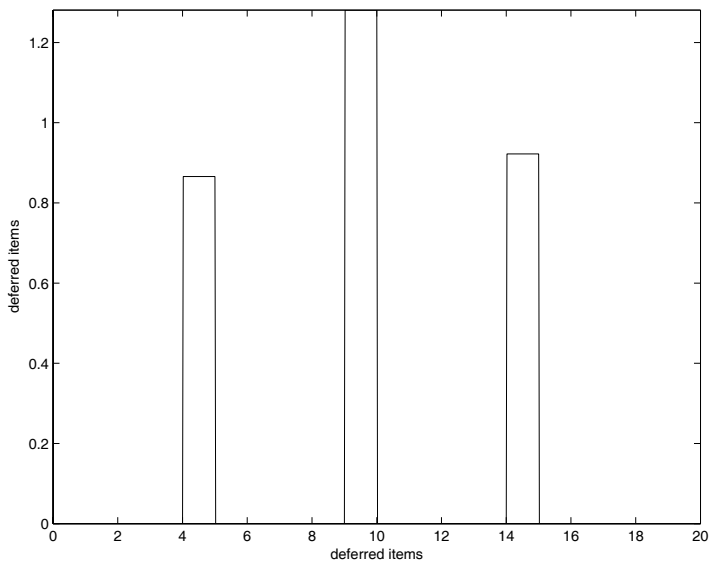


Figure 12.9. Stock level and demand (stochastic case)

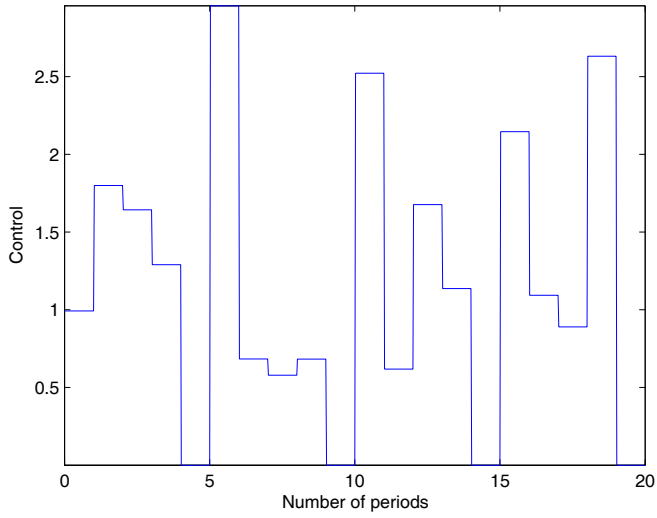


Figure 12.10. Production rate (stochastic case)

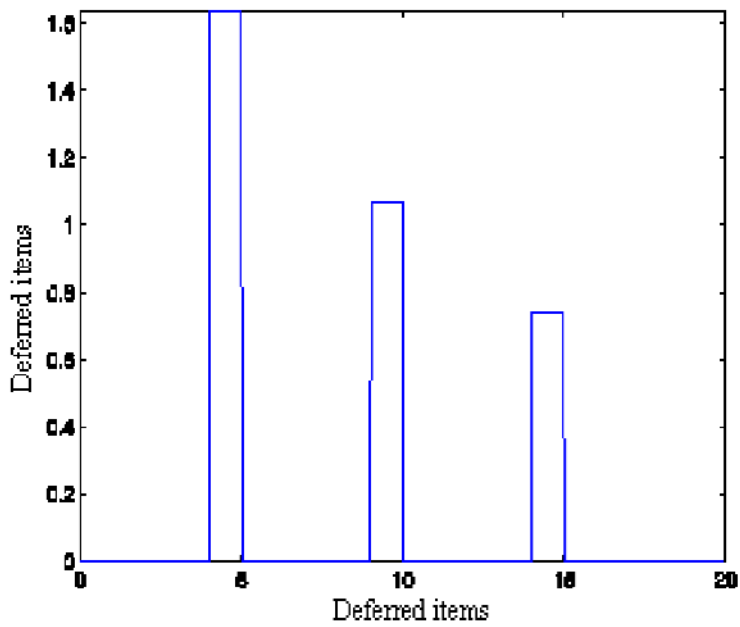


Figure 12.11. Deferred items (stochastic case)

- The optimization problem at level two becomes

$$P3' \left\{ \begin{array}{l} \min \sum_{k=1}^N [c^x x(k) + c^u u(k) + c^{w_1} w_1(k) + c^{w_2} w_2(k) + c^{w_3} w_3(k)] \\ \text{s.t. :} \\ x(k) = x(k-1) + u(k) + w_1(k) - w_1(k-1) + w_2(k) - w_2(k-2) \\ \quad + w_3(k) - w_3(k-3) - d(k), x(0) = x_0 \\ u(k) \leq \bar{u} \\ u(k) \geq 0 \\ x(k) \geq 0 \end{array} \right. \quad (12.48)$$

With the same data of Table 12.1 solving the optimization problems at the two levels, we get the results illustrated by Figures 12.12–12.18. Figure 12.12 gives the same results as for the case without deferred items. Figures 12.19–12.21 give the stock level and the production at different periods.

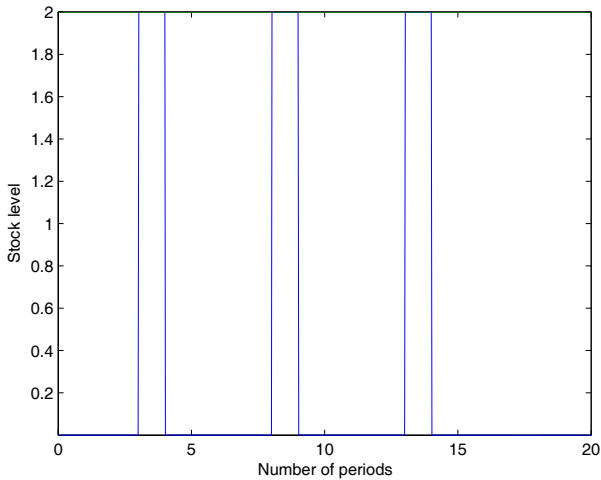


Figure 12.12. Stock level and demand (deterministic case)

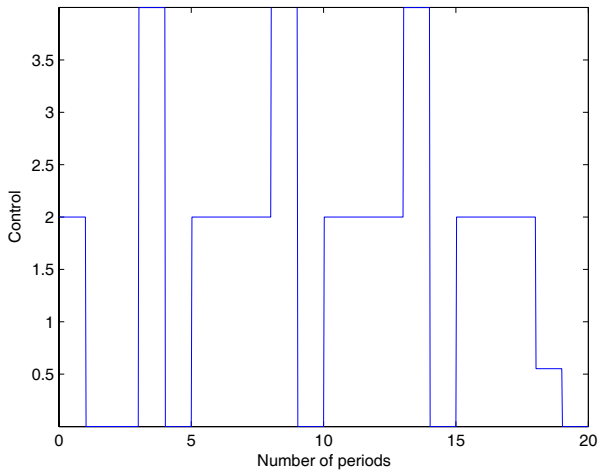


Figure 12.13. Production rate (deterministic case)

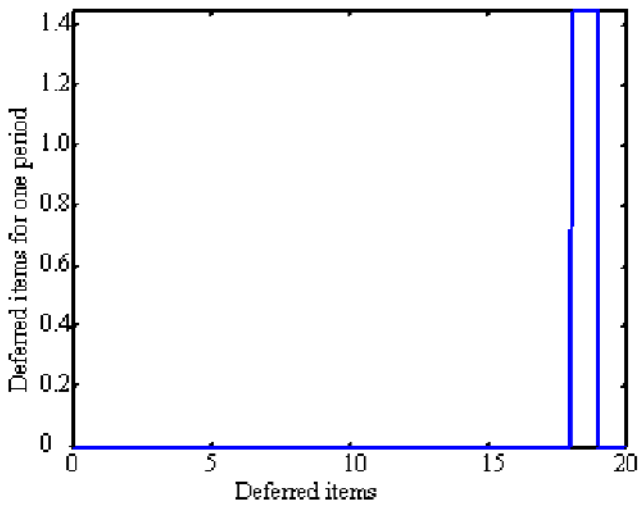


Figure 12.14. Deferred items for one period (deterministic case)

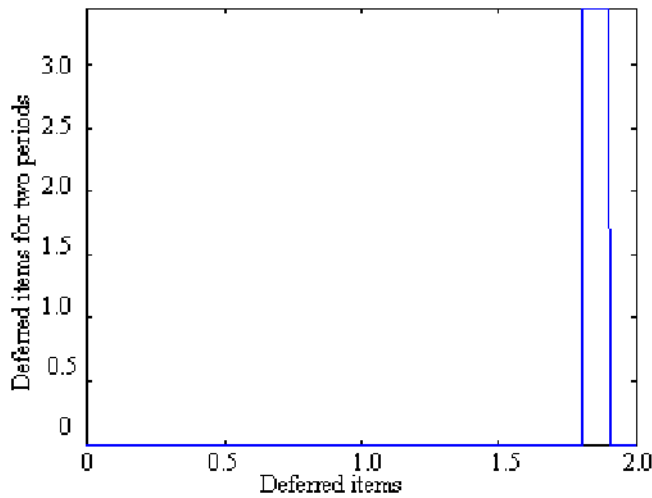


Figure 12.15. Deferred items for two period (deterministic case)

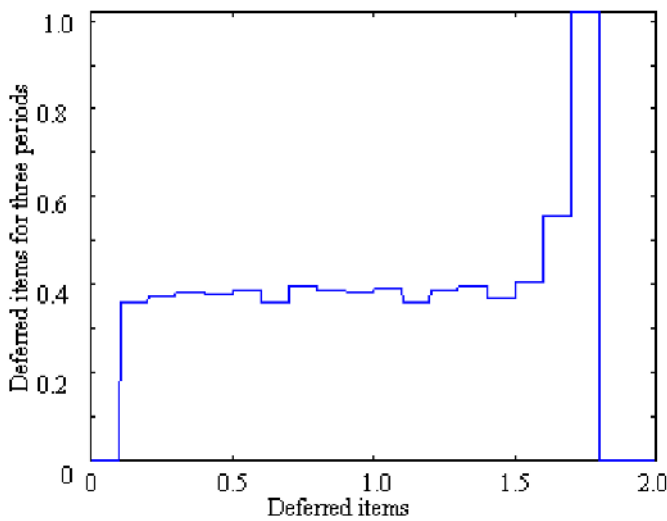


Figure 12.16. Deferred items for three period (deterministic case)

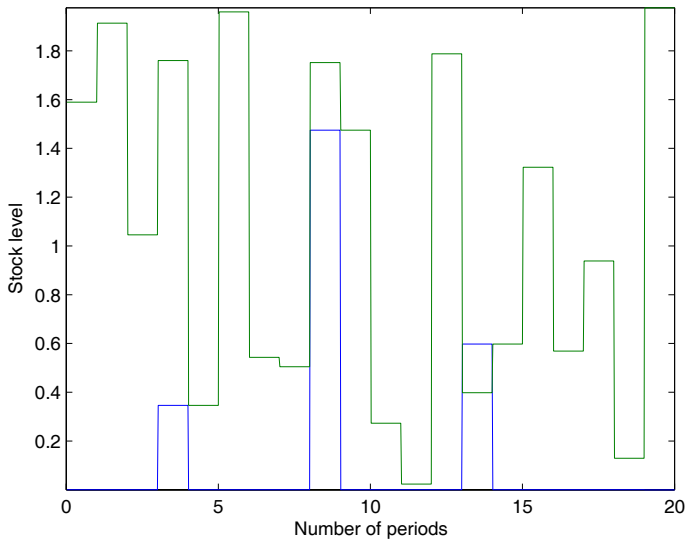


Figure 12.17. Production rate (stochastic case)

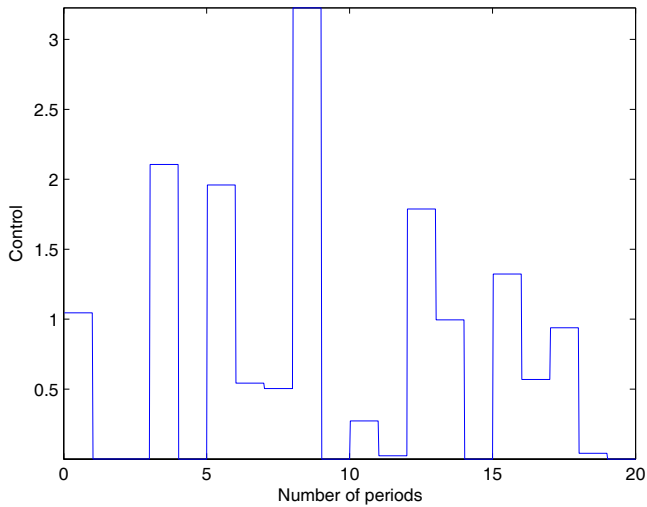


Figure 12.18. Stock level and demand (stochastic case)

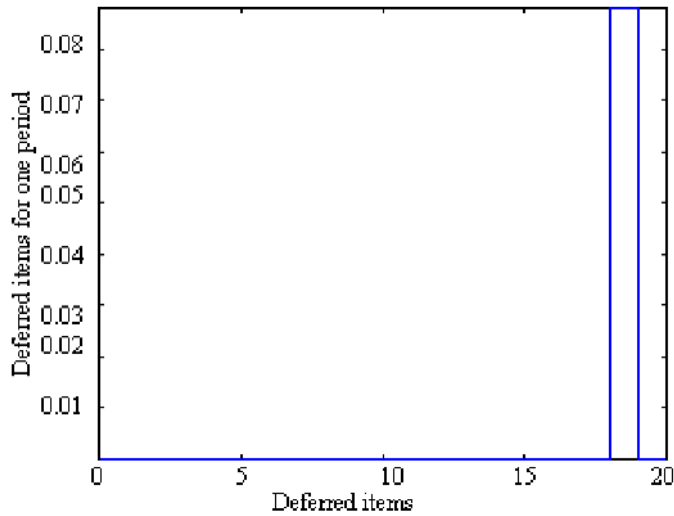


Figure 12.19. Deferred items for one period (stochastic case)

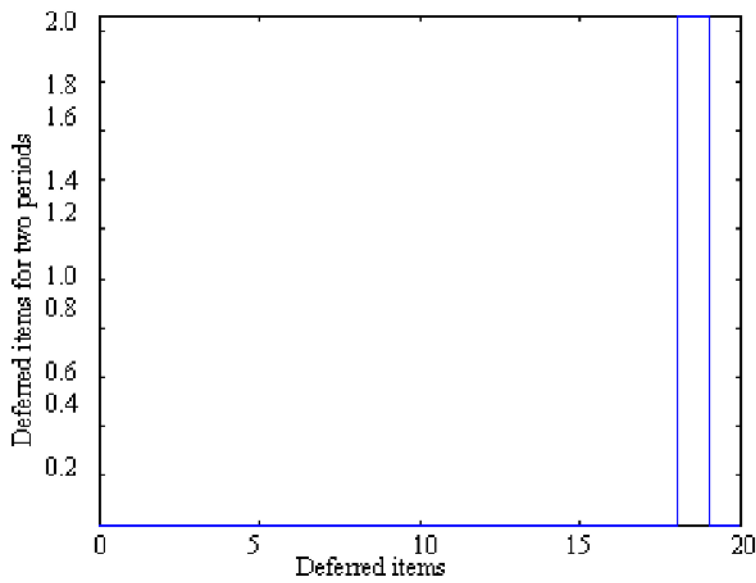


Figure 12.20. Deferred items for two period (stochastic case)

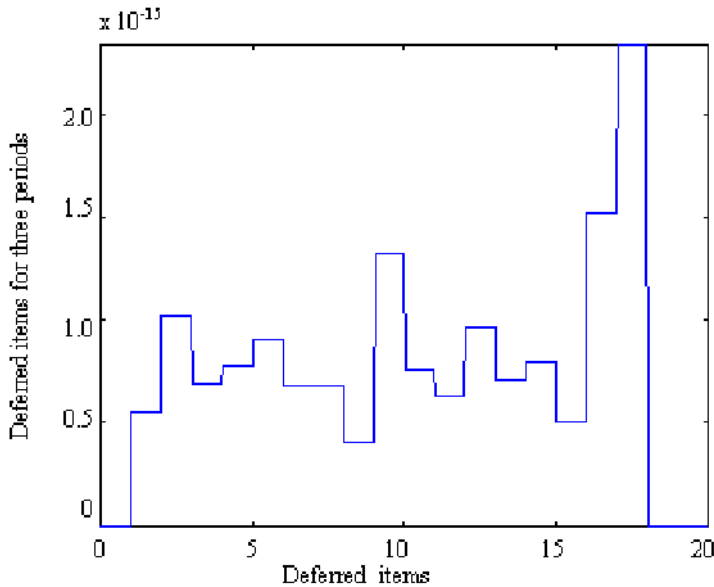


Figure 12.21. Deferred items for three period (stochastic case)

We can make more extensions for our model to include the following facts:

- Model with depreciation;
- Model with depreciation after some periods of time; and
- Model with setups.

12.5 Conclusion

In this chapter we have tackled the production and preventive maintenance control problem for manufacturing system with random breakdowns. This problem is formulated as a stochastic optimal control problem where the state of the production system is modeled as a Markov chain, the demand is constant and the produced items are assumed to deteriorate with a given rate γ . With some assumptions, the optimal production rate is still hedging point policy with some changes at the hedging point x^* . The production and preventive maintenance problem has also been solved using a hierarchical approach with two levels. The level one determines the instants when the maintenance has to be performed. The level two determines the production to track the demand. Some extensions of this model have been proposed.

References

- Aghezzaf Jamali MA, Ait-Kadi D (2007) An Integrated Production and Preventive maintenance Planning Model. *Eur J of Oper Res* 181: 679–685.
- Akella R, Kumar PR (1986) Optimal Control of Production Rate in a Failure Prone Manufacturing System. *IEEE Trans on Autom Control* AC-31(2):106–126.
- Algoet PH (1988) Flow balance equations for the steady-state distribution of a flexible manufacturing system. *IEEE Trans on Autom Control* 34(8): 917–921.
- Bertsekas DP (1987) *Dynamic Programming: Deterministic and Stochastic Models*. Prentice Hall Inc.
- Bielecki T, Kumar PR (1988) Optimality of zero-inventory policies for unreliable manufacturing systems. *Oper Res* 36(4): 532–541.
- Boukas EK (1987) *Commande optimale stochastique appliquée aux systèmes de production*, Ph.D. Thesis, Ecole Polytechnique de Montréal, Université de Montréal.
- Boukas EK (1993) Control of Systems with Controlled Jump Markov Disturbances. *Control Theory and Adv Technol* 9(2): 577–595.
- Boukas EK (1995) Numerical Methods for HJB Equations of Optimization Problems for Piecewise Deterministic Systems. *Optima Control Appl & Methods* 16: 41–58.
- Boukas EK (1998) Hedging point policy improvement. *J of Optim Theory and Appl* 97(1).
- Boukas EK (2006) *Manufacturing Systems: LMI Approach*. *IEEE Trans on Autom Control* 51(6): 1014–1018.
- Boukas EK, Haurie A (1988) Optimality conditions for continuous time systems with controlled jump Markov disturbances: Application to production and maintenance scheduling. *Proceeding INRIA 8th International Conference Springer Verlag on Analysis and Optimization of Systems*, Antibes.
- Boukas EK, Haurie A (1990) Manufacturing flow control and preventive maintenance: A stochastic control approach. *IEEE Trans on Autom Control* 35(9): 1024–1031.
- Boukas EK, Kenne JP (1997) Maintenance and production control of manufacturing system with setups. *Lect in Appl Math* 33: 55–70.
- Boukas EK, Liu ZK (2003) Modeling and Numerical Methods in Manufacturing Systems using Control Theory. In: Shanbhag DN, Rao CR (eds.) *Handbook of Statistics 21: Stochastic Processes: Modeling and Simulation*.
- Boukas EK, Yang H (1996) Manufacturing Flow Control and Preventive Maintenance: A Stochastic Control Approach. *IEEE Transaction on Automatic Control* 41(6): 881–885.
- Boukas EK, Zhu Q, Zhang Q (1994) Piecewise deterministic Markov process model for flexible manufacturing systems with preventive maintenance. *J of Optim Theory and Appl* 81(2): 258–275.
- Boukas EK, Zhang Q, Yin G (1996) On robust design for a class of failure prone manufacturing system. In: Yin, Zhang (eds.) *Recent advances in control and optimization of manufacturing systems*. *Lecture Notes in control and information sciences* 214, Springer-Verlag, London.
- Bremaud P, Malhamé RP, Massoulié L (1997) A manufacturing system with general failure process: stability and IPA of hedging control policies. *IEEE Trans on Autom Control* 42(2): 155–170.
- Buzacott JA, Shanthikumar JG (1993) *Stochastic Models of Manufacturing Systems*. Prentice-Hall, Englewood Cliffs, NJ.
- Caramanis M, Liberopoulos G (1992) Perturbation analysis for the design of flexible manufacturing system flow controllers. *Oper Res* 40: 1107–1125.
- Connolly S, Dallery Y, Gershwin SB (1992) A real-time policy for performing setup changes in a manufacturing system. *Proceeding of the 31st IEEE Conference on Decision and Control*, Tucson, AZ.

- Costa OLV (1989) Average impulse control of piecewise deterministic processes. *IMA J of Math Control and Inf* 6(4): 379–397.
- Costa OLV (1991) Impulse control of piecewise-deterministic processes via linear programming. *IEEE Trans on Autom Control* 36(3): 371–375.
- Costa OLV, Davis MHA (1989) Impulse control of piecewise-deterministic processes. *Math of Control, Signals, and Syst* 2(3): 187–206.
- Dassios A, Embrechts P (1989) Martingales and insurance risk. *Commun Stat Stoch Models* 5(2): 181–217.
- Davis MHA (1984) Piecewise deterministic Markov processes: A general class off non-diffusion stochastic models. *J of Royal Stat Soc* 46(3): 353–388.
- Davis MHA (1986) Control of piecewise-deterministic processes via discrete-time dynamic programming. In: Helmes K, Christopheit, Kohlman (eds.) *Stochastic Differential Systems*. Springer-Verlag, pp 140–150.
- Davis MHA (1993) *Markov Modeling and Optimization*. Chapman and Hall.
- Davis MHA, Dempster MAH, Sethi SP, Vermes D (1987) Optimal capacity expansion under uncertainty. *Adv in Appl Probab* 19: 156–176.
- Dempster MAH (1991) Optimal control of piecewise deterministic Markov processes. In: *Applied Stochastic Analysis*. Gordon and Breach, New York, pp 303–325.
- Desrochers AA (1990) *Modeling and Control of Automated Manufacturing Systems*. IEEE Computer Society Press, Washington.
- Duncan TE, Pasik-Duncan B, Zhang Q (1999) Adaptive control of stochastic manufacturing system with hidden Markovian demands and small noise. *IEEE Trans on Autom Control* 44(2): 427–431.
- Fleming WH, Soner HM (1992) *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, New York.
- Fu K, Hsu VN, Lee CY (2006) Inventory and Production Decisions for an Assemble-to-Order System with Uncertain Demand and Limited Assembly Capacity. *Oper Res* 54(6): 1137–1150.
- Gatarek D (1992) Optimality conditions for impulsive control of piecewise-deterministic processes. *Math of Control, Signals, and Syst* 5(2): 217–232.
- Gershwin SB (1989) Hierarchical flow control: A framework for scheduling and planning discrete events in manufacturing systems. *Proc of the IEEE, Special Issue on Discret Event Dyn Syst* 77(1): 195–209.
- Gershwin SB (1993) *Manufacturing Systems Engineering*. Prentice Hall, Englewood Cliffs.
- Gershwin SB, Caramanis M, Murray P (1988) Simulation experience with a hierarchical scheduling policy for a simple manufacturing system. *Proceeding of the 27th IEEE Conference on Decision and Control*, Austin, TX, 1941–1849.
- Gonzales R, Roffman E (1985) On the deterministic control problems: an approximation procedure for the optimal cost I. The stationary problem. *SIAM Control and Optim* 23(2): 242–266.
- Haurie A, L'Ecuyer P (1986) Approximation and bounds in discrete event dynamic programming. *IEEE Trans on Autom Control* 31(3): 227–235.
- Hu JQ, Caramanis M (1994) Dynamic set-up scheduling of flexible manufacturing systems: Design and stability of near optimal general round robin policies. In: Kumar PR, Varaiya P (eds.) *Discrete Event Systems*, IMA volumes in mathematics and applications series, Springer-Verlag.
- Hu JQ, Xiang D (1994) Structural properties of optimal flow control for failure prone production system. *IEEE Trans on Autom Control* 39(3): 640–642.
- Hu JQ, Vakili P, Yu GX (1994) Optimality of hedging point policies in the production control of failure prone manufacturing systems. *IEEE Trans on Autom Control* 39(9): 1875–1880.

- Kimemia J, Gershwin SB (1983) An algorithm for computer control of a flexible manufacturing system. *IIE Trans* 15(4): 353–362.
- Krasovskii NN, Lidskii EA (1961) Design of Controller in System with Random Attributes-Part 1. *Automa Remote Control* 22: 1021–1025.
- Kushner HJ (1977) *Probability Methods for Approximation in Stochastic Control and for Elliptic Equations*. Academic Press, New York.
- Kushner HJ, Dupuis PG (1992) *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag, New York.
- Lejeune MA, Ruszczynski A (2007) An Efficient Trajectory Methods for Probabilistic Production Inventory Distribution Problems. *Oper Res* 55(2): 378–394.
- Lenhart MS, Liao YC (1988) Switching control of piecewise-deterministic processes. *J of Optim Theory and Appl* 59(1): 99–115.
- Liberopoulos G, Caramanis M (1994) Production control of manufacturing system with production rate-dependent failure rates. *IEEE Trans on Autom Control* 39(4): 889–895.
- Maimoun O, Khenelnitsky E, Kogan K (1998) *Optimal flow control in manufacturing system: Production Planning and Scheduling*. Kluwer Academic Publishers, Dordrecht.
- Marquez AC, Gupta JND, Heguedas AS (2007) Maintenance Policies for a Production System with Constrained Production Rate and Buffer Capacity. *Int J of Prod Res* 41(9): 1909–1926.
- Menaldi JL (1989) Some estimates for finite difference approximations. *SIAM J on Control and Optim* 27: 579–607.
- Olsder GJ, Suri R (1980) Time optimal of parts-routing in a manufacturing system with failure prone machines. *Proc. 19th IEEE. Conference on Decision and Control Albuquerque, New Mexico*, 722–727.
- Panogiotidou S, Tagaras G (2007) Optimal Preventive Maintenance for Equipment with two Quality States and General Failure Time Distributions. *Eur J of Oper Res* 180: 329–353.
- Rishel R (1975) Control of systems with jump Markov disturbances. *IEEE Trans on Autom Control* 20: 241–244.
- Ross SM (1970) *Applied Probability Methods with Optimization Applications*. Holden-Day, San Francisco.
- Sethi SP, Zhang Q (1994) *Hierarchical Decision Making in Stochastic Manufacturing Systems*. Birkhauser, Boston.
- Sethi SP, Taksar MI, Zhang Q (1997) Optimal production planning in a stochastic manufacturing system with long-run run average cost. *J of Optim Theory and Appl* 92(1): 161–188.
- Sharifnia A (1988) Production control of a manufacturing system with multiple machine states. *IEEE Trans on Autom Control* 33(7): 620–625.
- Sharifnia A, Caramanis M, Gershwin SB (1991) Dynamic setup scheduling and flow control in manufacturing systems. *Discret Event Dyn Syst: Theory and Appl* 1: 149–175.
- Soner HM (1986) Optimal control with state-space constraint II. *SIAM J on Control and Optim* 24(6): 1110–1122.
- Sworder DD, Robinson VG (1974) Feedback regulators for jump parameter systems with state and control dependent transition rates. *IEEE Trans on Autom Control* 18: 355–359.
- Tu FS, Song DP, Lou SXC (1993) Preventive hedging point control policy and its realization, Preprint of the XII World IFAC congress, July 18–23, Sydney, Australia, 5: 13–16.
- Veatch MH, Caramanis MC (1999) Optimal average cost manufacturing flow controllers: convexity and differentiability. *IEEE Trans on Autom Control* 44(4): 779–783.

- Viswanadham N, Narahari Y (1992) Performance modeling of automated manufacturing systems. Prentice Hall, Englewood Cliffs.
- Wang H (2002) A Survey of Maintenance Policies of Deteriorating Systems. *Eur J of Oper Res* 139: 469–489.
- Wonham WM (1971) Random differential equations in control theory. In: Bharucha-Reid AT (ed.) *Probabilistic Methods in Applied Mathematics*, Vol. 2. Academic, New York.
- Yan H, Zhang Q (1997) A numerical method in optimal production and setup scheduling of stochastic manufacturing system. *IEEE Trans on Autom Control* 42(10):1452–1555.
- Yang J, Qi X, Xia Y (2005) A Production Inventory System with Markovian Capacity and Outsourcing Option. *Oper Res* 53(2): 328–349.
- Yin G, Zhang Q (1996) Recent Advances in Control and Optimization of Manufacturing System. *Lecture notes in control and information sciences* 214, Springer-Verlag, London.
- Yin G, Zhang Q (1998) *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*. Springer-Verlag.
- Yu XZ, Song WZ (1999) Further properties of optimal hedging points in a class of manufacturing system. *IEEE Trans on Autom Control* 44(2): 379–382.

Maintenance Strategies

Inspection Strategies for Randomly Failing Systems

Anis Chelbi and Daoud Ait-Kadi

13.1 Introduction

In many situations there are no apparent symptoms indicating the imminence of failure. For such systems whose failures are not self-announcing, the level of degradation can be known only through inspection. Detection and alarm systems as well as stand-by systems are some examples of such equipment which must be inspected. Each inspection consists in measuring one or some characteristics to assert the degradation level. An inspection strategy establishes the instants at which one or more operating parameters have to be controlled, in order to determine if the system is in an operating or a failure state. These inspections require human and material resources as well as a certain know how.

Given that failure can be detected only following an inspection, the system remains in a failed state between the instant of failure occurrence and the instant of its detection. This inactivity period might cause significant losses. Hence, it is crucial to determine the sequence of inspection instants which optimizes a certain performance criterion over a given time span. Generally, we look for the inspections sequence minimizing the total average cost per time unit or maximizing the system steady state availability.

The first scientific works which have been devoted to optimal inspection policies for randomly failing systems, are those of Savage (1956), Barlow *et al.* (1960), Derman (1961), Coleman and Abrams (1962), Noonan and Fain (1962), and Weiss (1962, 1963).

A lot of works have been recently published on the inspection problem. The proposed models may be classified according to the system's operating context such as the maintenance actions, the quality and the quantity of information available, the performance criteria, the time span and all constraints related to operating conditions and resources availability.

Basically, two general situations are considered in the literature. The first consists in a black-box approach with binary states associated to the equipment (working or failed). Inspections consist simply in assessing if the equipment is working or in a failed state. The second approach deals with situations where it is possible, through direct or indirect control, to assess the equipment condition and

eventually take preventive measures before failure occurrence. This approach is commonly applied in condition-based maintenance (C.B.M.).

The content of this chapter is organized as follows. In Section 13.1, the first fundamental contribution is presented in detail. The modeling approach as well as numerical procedures to generate optimal inspection schedules will be highlighted. Section 13.2 is dedicated to the extensions of the basic model, mainly those addressing the frequent inspections case, the situations where the system lifetime distribution is unknown, as well as the case where inspections affect the equipment state, models where system availability is taken as the performance criterion instead of the cost, and those focusing on systems which alternate between periods of activity and periods of inactivity. In Section 13.3, inspection policies for multi-component systems will be presented. Models will be grouped in sub-sections according to the following sequence: models based on the failure tree method, those which consider the cases of cold and hot stand-by systems with known and partially known lifetime distributions, and finally those dealing with the case of systems with components failure dependency. All these models assume that the equipment is replaced by a new and identical one if the inspection reveals that it is in a failed state; otherwise no action is taken. Hence, replacement occurs only if failure is detected. In Section 13.4, we will expose the essential parts of inspection models developed in the literature in the context of condition-based maintenance for both single- and multi-component systems. Concluding remarks and some further potential research will be presented. Two tables presenting a classification of the considered references are provided at the end of this chapter (Table 13.1 and Table 13.2).

13.1.1 Notation

The following set of notations will be used throughout this chapter:

- $f(t)$: probability density function associated with the equipment lifetime;
- $F(t)$: probability distribution function associated with the equipment lifetime;
- $R(t)$: equipment reliability function;
- $r(t)$: equipment instantaneous failure rate function;
- μ : equipment average lifetime;
- Y : equipment state variable;
- UTR: equipment stationary availability (Up Time Ratio);
- $A(t)$: equipment instantaneous availability;
- C_i : constant cost associated with each inspection;
- C_d : constant cost incurred for each time unit of inactivity between failure and its detection;

C_g : constant cost of replacing the system by a new identical one once failure is detected following inspection;

C_p : average cost of each preventive maintenance action;

C_f : average cost incurred following a false alarm;

C_o : cost of operation per unit time;

C_r : replacement cost of a failed unit;

$X=(x_1, x_2, \dots)$: inspection times sequence;

$C(X)$: total average cost during a replacement cycle when inspection are performed according to the inspection sequence X ;

$T(X)$: average replacement cycle duration following the sequence X ;

$R_c(.)$: total expected cost per unit time over an infinite span associated with the inspection strategy;

τ : inspection period;

T : equipment age at which inspection must be performed

$n(t)$: continuous function expressing the number of inspections per time unit;

λ : equipment constant failure rate;

$(1-p)$: probability of failure detection following an inspection;

q : probability of non-self-announcing failure (probability to be in an idle period according to mission profile);

$\delta : 1/(1-\sigma)$ with σ standing for the probability of undetected failure following inspection;

α : probability of having a false alarm following inspection;

T_a : mean duration of a preventive maintenance action;

T_f : average down-time of the system due to a false alarm;

T_c : mean duration of a corrective maintenance action;

T_i : mean duration of an inspection;

L : control parameter threshold level.

13.2 Basic Inspection Model

13.2.1 Problem Definition

Consider a non-self-announcing-failure equipment inspected at instants $x_1, x_2, x_3 \dots$ (see Figure 13.1). when inspection reveals that the equipment is in a failed state, it is immediately replaced by a new identical one (or restored to a state as good as new). When inspection shows that it is still in a good state, no action is undertaken and the equipment remains in service.

Costs are associated with inspection, inactivity and replacement. The objective is to find the optimal inspection instants x_i ($i = 1, 2, \dots$) which minimize the total average cost per time unit over a given horizon.

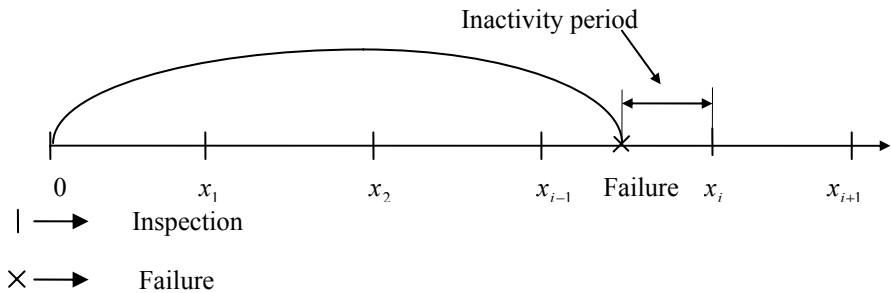


Figure 13.1. The sequence of inspection instants

13.2.2 Working Assumptions and Mathematical Model

The following assumptions are specifically made:

1. The equipment is either in an operating or a failed state;
2. Failure can be known only through inspection;
3. Inspections have negligible durations;
4. Inspection does not affect the equipment state;
5. Inspection reveals the right state of the equipment with certainty (perfect inspections);
6. A constant average cost is associated with each inspection;
7. A constant average cost is incurred for each time unit of inactivity between failure and its detection;
8. In case of failure, replacement is perfectly performed; and
9. The inspection process ends once failure is revealed.

If failure occurs at instant t between the k th and the $(k+1)$ th inspections, then the average cost would be the sum of costs related to the $(k+1)$ inspections performed and to the $(x_{k+1} - t)$ time units of inactivity (see Figure 13.2, Barlow *et al.*, 1963 and Barlow and Proschan, 1965).

As failure may occur within any time interval $[x_k, x_{k+1}]$, $k = 0, 1, 2, \dots$, the total average cost is expressed as follows:

$$C(x_1, x_2, \dots) = \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} [C_i (k+1) + C_d (x_{k+1} - t)] f(t) dt + C_r \quad (13.1)$$

with $x_0 = 0$.

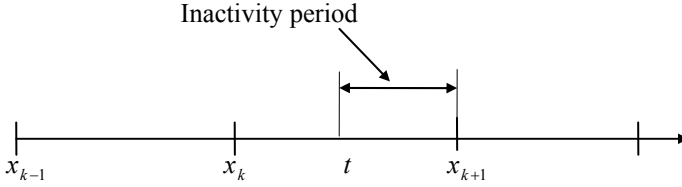


Figure 13.2. Inactivity time

The objective is to find the optimal inspection sequence x_k ($k = 1, 2, \dots$) which minimize $C(x_1, x_2, \dots)$. Barlow and Proschan (1965) have suggested a numerical procedure to generate those instants in the case of a particular class of probability density functions called ‘*Polya frequency functions of order 2*’, which represents a generalization of functions with increasing failure rate. The following algorithm is used by the authors to compute the optimum inspection schedule:

Step 1– select x_1 that satisfy $C_i = C_d \int_0^{x_1} (x_1 - t) f(t) dt$

Step 2– compute recursively the computing to obtain $x_1, x_2, \dots$ from the following relationship:

$$x_{k+1} - x_k = \frac{F(x_k) - F(x_{k-1})}{f(x_k)} - \frac{C_i}{C_d}$$

Step 3– if any $\delta_k > \delta_{k-1}$, reduce x_1 and repeat, where $\delta_k \equiv x_{k+1} - x_k$. If any $\delta_k < 0$, increase x_1 and repeat; and

Step 4– continue until $x_1 < x_2 < \dots$ to obtain the optimal inspection sequence.

This procedure turned out to be quite cumbersome, especially due to its iterative nature and the difficulty to choose an appropriate value of the first inspection instant x_1 .

In a second model, Barlow and Proschan (1965) considered the same problem described above. Assumptions (1) to (8) still hold but the last assumption (9) is replaced by: once failure is detected, the system is repaired or replaced by a new identical one incurring a constant average cost C_r , r time units are necessary to perform this repair or replacement. The system is then considered as good as new and the inspection procedure starts again. In such situations, the optimal inspection policy is the one which minimizes the total average cost per time unit over an infinite horizon.

If inspections are performed at instants $x_1 < x_2 < \dots$, then the total average cost per time unit over an infinite horizon is

$$R_c = \frac{C(X)}{T(X)} \quad (13.2)$$

$C(x_1, x_2, \dots)$ is given by the following expression:

$$C(x_1, x_2, \dots) = \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} [C_i (k+1) + C_d (x_{k+1} - t)] f(t) dt + C_r \quad (13.3)$$

and $T(x_1, x_2, \dots)$ is expressed as follows:

$$T(x_1, x_2, \dots) = \mu + \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} (x_{k+1} - t) f(t) dt + r \quad (13.4)$$

A second numerical procedure has been suggested to generate the optimal inspection sequence $X^* = x_1^*, x_2^*, \dots, x_n^*$

In conclusion, for the case of both algorithms shown above, the main problem resides in the choice of the first inspection instant x_1 in order to obtain the desired degree of precision.

13.3 Extensions of the Basic Model

13.3.1 Inspection Models for Single Component Systems

13.3.1.1 Nearly Optimal Inspection Sequences

Nakagawa and Yasui (1980) and Nakagawa (2005), have reconsidered the optimal strategy proposed by Barlow and Proschan (1965). They proposed a new procedure to generate a nearly-optimal strategy, much easier to obtain, allowing one to calculate the inspection instants backwards from an instant x_n quite distant in time.

For situations where the process ends once failure is detected, the following algorithm is proposed:

Step 1—choose a real number ε in the interval $\left[0, \frac{C_i}{C_d}\right]$;

Step 2—choose an inspection instant x_n quite distant to have a good precision;

Step 3—calculate x_{n-1} to satisfy

$$x_n - x_{n-1} - \varepsilon = \frac{F(x_n) - F(x_{n-1})}{f(x_n)} - \frac{C_i}{C_d} \quad (13.5)$$

Step 4—calculate $x_{n-1} > x_{n-2}$ recursively using Equation 13.5;

Step 5—continue until one of the two following conditions is satisfied:

$$x_k < 0 \text{ or } x_{k+1} - x_k > x_k .$$

This algorithm is based on the same approach as the one proposed by Barlow and Proschan (1965) with the difference that its execution doesn't give way to any numerical divergences.

The nearly optimal policy generated with the procedure of Nakagawa and Yasui (1980) has been compared to the optimal strategy in the case of a Weibull distribution:

$$F(t) = 1 - e^{-(\lambda t)^\alpha} \quad (13.6)$$

with $\lambda = 0.002$ and $\alpha = 2$, the procedure provides a very good approximation of the optimal strategy, particularly when $\varepsilon = 4.5$ and $\frac{C_i}{C_d} = 10$

The performance of the procedure still depends on the choice of ε . Nakagawa and Yasui (1980) suggest the value of ε such that $\varepsilon = \frac{1}{2} \left(\frac{C_i}{C_d} \right)$, and very good results are obtained.

This nearly-optimal strategy could be used to find an initial approximation of the first inspection instant x_1 , required to determine the optimal sequence using the algorithm of Barlow and Proschan (1965).

In order to overcome the difficulty of having a problem with n variables, $x_1, x_2, \dots, x_n$, Munford and Shahani (1972) have proposed an algorithm based on a single parameter to generate a nearly-optimal inspection sequence. The authors present an asymptotic method to obtain the optimal inspection sequence and assume that the probability that a unit with age x_{k-1} fails in an interval $(x_{k-1}, x_k]$ is constant for all k :

$$\frac{F(x_k) - F(x_{k-1})}{\bar{F}(x_{k-1})} \equiv p \quad \text{where } (k=1, 2, \dots)$$

noting that $F(x_1) = p$.

The equation given above can be solved for x_k , and we obtain

$$\bar{F}(x_k) = q^k \quad \text{or} \quad x_k = \bar{F}^{-1}(q^k) \quad \text{where } (k=1, 2, \dots)$$

where $q \equiv 1 - p$ ($0 < p < 1$); from Equation 13.3 the total expected cost is expressed as follows

$$C(p) = \frac{C_i}{p} + C_d \sum_{k=1}^{\infty} x_k q^{k-1} p - C_d \mu + C_r$$

The objective is to find the p that minimizes $C(p)$, Moreover they showed that their algorithm has the property to generate decreasing, constant or increasing inspection sequences, if the considered system has a decreasing, constant or increasing failure rate.

13.3.1.2 Case of Frequent Inspections

Keller (1974) noticed that the optimal strategy (Barlow and Proschan, 1965) becomes more complex when inspections instants are very frequent. In this particular case, he supposed that the inspection process can be described by a

continuous intensity function $n(t)$ expressing the number of inspections per time unit:

$$\int_0^{x_n} n(t) dt = n \quad (13.7)$$

It is assumed that the mean time from failure at time t to its detection at time $t + a$ is half of a checking interval:

$$\int_t^{t+a} n(u) du = \frac{1}{2}$$

This expression is approximated as follows:

$$\int_t^{t+a} n(u) du \approx \frac{1}{2} an(t)$$

Thus, $a = \frac{1}{[2n(t)]}$ and the inspections are planned according to a period $1/n(t)$.

The total expected cost is given by

$$\begin{aligned} C(n(t)) &= \int_0^\infty \left[C_i \int_0^t n(u) du + \frac{C_d}{2n(t)} \right] dF(t) + C_r \\ &= \int_0^\infty \bar{F}(t) \left[C_i n(t) + \frac{C_d r(t)}{2n(t)} \right] dF(t) + C_r \end{aligned}$$

An explicit solution is obtained in the case where the cost of loss per time unit is constant in differentiating $C(n(t))$ with $n(t)$ and putting to zero. The optimal solution $n(t)$ is given by

$$n(t) = \left\{ \frac{C_d r(t)}{2C_i} \right\}^{1/2} \quad (13.8)$$

The optimum inspection time is obtained by the following expression:

$$n = \int_0^{x_n} \sqrt{\frac{C_d}{2C_i} r(t)} dt \quad (n=1,2,3,\dots)$$

Notice that in this case, the function $n(t)$ is proportional to the square root of the system's failure rate. These results have been applied to systems with a constant failure rate λ , the inspection period is given by

$$\frac{1}{n(t)} = \left(\frac{\lambda C_d}{2C_i} \right)^{-1/2} \quad (13.9)$$

This result is in accordance with the exact solution in the case of a periodic inspection process, which is valid and justified for systems with constant failure

rate. Let's point out here that periodic inspections are much used in practice in various fields, like for example statistical quality control (Taguchi *et al.* 1989), medicine, nuclear energy, defence, *etc.* Many authors like Rodrigues (1983, 1990) and Nakagawa and Yasui (1979) have worked on this problem which consists mainly in finding the optimal inspection period.

Kaio and Osaki (1984) extend the Keller's (1974) model, and present an algorithm which generates a nearly-optimal inspection sequence. According to their model, the inspection instants x_n satisfy Equation 13.7.

A nearly-optimal inspection sequence is obtained by substituting $n(t)$ in Equation 13.7 by Keller's expression given by Equation 13.8.

Kaio and Osaki (1989) compared the algorithm of Barlow and Proschan (1965) which generates optimal sequences to those of Munford and Shahani (1972), Nakagawa and Yasui (1980) and Kaio and Osaki (1984) which generate quasi-optimal sequences. They made a comparison in the cases of Gamma and Weibull lifetime distributions. They concluded that there is no significant difference between the optimal sequence and the three nearly-optimal sequences in both cases. However, they recommended the algorithm of Kaio and Osaki, first because of its simplicity, second because it presents absolutely no restriction with respect to lifetime distributions, and third because it can incorporate more complex inspection policies.

13.3.1.3 Case of Unknown System Lifetime Distribution

In the case where the system lifetime distribution is unknown, Leung (2001) studied the four situations described below. He developed, for each one, an optimal inspection policy based on the Keller's expression of the number of inspections per time unit $n(t)$ (Equation 13.7), considering a finite time span $[0, T]$. He obtains the optimal inspection sequence by combining, for each of the four situations, Equation 13.7 with the following equations:

- Situation 1: basic model corresponding to the same assumptions as Keller's:

$$n(t) = \frac{1}{2} \sqrt{\frac{C_d \beta}{C_i (1 - \beta t)}} \quad \text{for } 0 \leq t \leq \frac{1}{\beta} \quad (13.10)$$

with $\beta = 1/T$

- Situation 2: basic model with imperfect inspections under the assumption that failure is detected with a constant probability equal to $(1 - p)$:

$$n(t) = \frac{1}{2} \sqrt{\frac{C_d (1 + p) \beta}{C_i (1 - p) (1 - \beta t)}} \quad \text{for } 0 \leq t \leq \frac{1}{\beta} \quad (13.11)$$

with $\beta = 1/T$

- Situation 3: basic model with a non-negligible inspection duration d_i :

$$n(t) = \frac{1}{2} \sqrt{\frac{C_d \beta}{C_i [1 - \beta(t - d_i)]}} \quad \text{for } 0 \leq t \leq \frac{1}{\beta} + d_i \quad (13.12)$$

with $\beta = 1/(T - d_i)$

- Situation 4: basic model with imperfect inspections and a non-negligible inspection duration d_i :

$$n(t) = \frac{1}{2} \sqrt{\frac{C_d(1+p)\beta}{C_i(1-p)[1-\beta(t-d_i)]}} \quad \text{for} \quad 0 \leq t \leq \frac{1}{\beta} + d_i \quad (13.13)$$

with $\beta = 1/(T - d_i)$

Still in the context of unknown equipment lifetime distribution, there is the work of Yang and Klutke (2000) who also propose a periodic inspection model.

13.3.1.4 Inspections Affecting the Equipment State

In many practical situations, in particular those related to industrial machinery, the inspection action might affect the state of the inspected equipment. Indeed, during inspection, each action performed by the operator may alter or improve the state of the equipment.

Thus, each time the equipment is inspected, its failure rate would be brought to a higher or lower level. These situations have been dealt with by Wattanapanom and Shaw (1978). They define a general expression of the failure rate which takes into account the system lifetime and the number of inspections already performed.

The system starts operating at instant t_0 with an *a priori* probability density function $f(t)$. If time is accelerated by a factor $\theta_k > 1$ following the k^{th} inspection, then the system, still operating at that instant t_k , would have been operating T_k time units on the original time scale. T_k is given by

$$T_k = t_1 + \theta_1(t_2 - t_1) + \dots + \theta_{k-1}(t_k - t_{k-1}) \quad (13.14)$$

This way, supposing that inspections stop at instant t_k , the lifetime conditional probability density function following that instant is expressed as follows:

$$f_k(t/t > t_k) = \frac{\theta_k \cdot f[T_k + \theta_k(t - t_k)]}{\int_{T_k}^{\infty} f(t)dt} \quad (13.15)$$

Equation 13.15 shows clearly that the lifetime conditional probability distribution depends on the number and the time distribution of preceding inspections.

Wattanapanom and Shaw (1978) managed to apply these algorithms only to systems with constant failure rates because the generalization of the problem requires the use of dynamic programming algorithms whose convergence is quite limited.

Chelbi and Ait-Kadi (1998) proposed a different approach with a strategy that takes into account the increase or decrease of the conditional probability of failure following each inspection. Their policy suggests that the system, having an increasing failure rate, is inspected at instants $(x_1, x_2, \dots)$ to determine if it is in an operating or in a failed state. They suppose that $F^{-1}(\cdot)$ exists and that inspections whose duration is negligible influence the degradation process of the system.

The problem is to determine the inspection sequence $(x_1, x_2, \dots)$ which minimizes the expected total cost per time unit, R_c , over an infinite horizon. This cost is expressed as follows:

$$R_c = \frac{E(C)}{E(T)} \quad (13.16)$$

where $E(C)$ stands for the average total cost for a replacement cycle whose duration is $E(T)$. $E(C)$ and $E(T)$ are respectively given by

$$E(C) = C_i \cdot E(I) + C_d \cdot E(A) + C_g \quad (13.17)$$

$$E(T) = \mu + E(A)$$

(13.18) where $E(I)$ represents the average number of inspections until failure detection, and $E(A)$ stands for the average inactivity duration. This model is based on the conditional probability p_i that failure occurs within the time interval $[x_{i-1}, x_i]$ given that the system was in an operating state at instant x_{i-1} :

$$p_i = \frac{F(x_i) - F(x_{i-1})}{1 - F(x_{i-1})} \quad \text{for } i = 1, 2, \dots \quad (13.19)$$

It is stated that in the case where the inspection alters the system's state and consequently accelerates its degradation process:

$$p_{i+1} > p_i \quad \text{for } i = 1, 2, \dots \quad (13.20)$$

Inversely, if the inspection reduces the failure rate or maintains it at its current level, then

$$p_{i+1} \leq p_i \quad \text{for } i = 1, 2, \dots \quad (13.21)$$

The inspection sequence can be obtained from Equation 13.19 in the following way:

$$x_i = F^{-1} \{ p_i [1 - F(x_{i-1})] + F(x_{i-1}) \} \quad \text{with } x_0 = 0 \quad (13.22)$$

In order to reduce the complexity of the generation of the optimal inspection sequence, the parameter p_i is expressed as a function of one unique parameter p_1 :

$$p_i = \Psi(p_1) \quad (13.23)$$

It should be noted that the model of Munford and Shahani (1972) cited above as well as that of Tadikamalla (1979) have considered such a function $\Psi(\cdot)$ as a constant to generate a nearly-optimal inspection sequence using the basic model of Barlow and Proschan (1965). The obtained results were close to the optimal solution.

13.3.1.5 System Availability as the Performance Criterion

In many practical situations, for security reasons, the criterion of cost may become of secondary importance; the system availability represents the primary performance criterion. The practitioners must maintain a certain level of availability with a minimum effort of inspection. This is particularly true for electronic alarm systems and many others. In general, such systems deteriorate because of the cumulated damage induced by transitory electromechanical shocks which occur randomly and have random magnitudes.

Wortman *et al.* (1994) have studied the problem for such systems; they proved that the equipment stationary availability is maximized when inspections follow a deterministic renewal process, which means that the interval between consecutive inspections remains constant. Inspections are performed independently of the shock process and according to a deterministic stationary renewal process with an average rate γ (inspections are carried out every $1/\gamma$ time units). Following each inspection, the system is replaced if inspection reveals that it is in failed state; otherwise, it remains in operation. The replacement duration is considered as negligible.

The authors established the expression of the system stationary availability, UTR, as a function of the inspection period $1/\gamma$ and the average number of shocks per time unit, ν , shocks occurring according to a Poisson process:

$$UTR = \frac{\gamma \cdot \left[\int_0^\infty R(z) \cdot g(z) dz + 1 \right]}{\nu \cdot \left[\sum_{n=0}^\infty \Phi\left(\frac{n}{\gamma}\right) \right]} \quad (13.24)$$

where $R(z)$ represents the expected number of shocks necessary to cumulate a total magnitude at least equal to z ; $R(z)$ is given by

$$R(z) = \sum_{k=1}^\infty S^k(z) \quad (13.25)$$

$S^k(\cdot)$ being the k th convolution of the shocks magnitude distribution function $S(\cdot)$ by itself.

n is the number of inspections, and $\Phi(\cdot)$ stands for the system survival function:

$$\Phi(t) = \int_0^\infty \sum_{k=0}^\infty S^k(z) \cdot \text{pois}(k, \nu, t) \cdot g(z) dz \quad (13.26)$$

where $\text{pois}(k, \nu, t)$ represents, for a Poisson process with mean νt , the probability to have k shocks within the time interval $[0, t]$.

Wortman *et al.* (1994) limited their study to situations where the times between consecutive shocks are exponentially distributed. Chelbi and Ait-Kadi (2000) generalized this model considering shocks distributed according to any given probability distribution $H(\cdot)$. The generalized expression of the system stationary availability becomes

$$UTR = \frac{\int_0^{\infty} \left[\sum_{k=1}^{\infty} (H^{(k)}(t) - H^{(k+1)}(t)) \int_0^{\infty} S^{(k)} g(z) dz \right] dt}{\left[\sum_{n=0}^{\infty} \Phi\left(\frac{n}{\gamma}\right) \right]} \quad (13.27)$$

Chelbi and Ait-Kadi (2000) present graphical results based on a numerical procedure, aiming at helping the decision makers to determine the inspection periods allowing reaching the required availability levels for a given set of input parameters.

In contrast to Wortman *et al.* (1994) and Chelbi and Ait-Kadi (2000) who worked on the systems stationary availability, Cui *et al.* (2004) studied the instantaneous availability of randomly failing systems submitted to periodic inspections every τ time units. They studied two models A and B. The first one considers that, following each inspection, the system is renewed even if it is not found in a failed state, whereas the second model states that if inspection reveals that the system is in an operating state, no action is undertaken and the system remains in the same state as before inspection. In addition, for each of the two models, two assumptions are alternatively taken into account, considering respectively renewal durations as a constant d or as a random variable following a probability density function $g(y)$ and a probability distribution $G(y)$.

For model A with constant renewal duration, the instantaneous availability is given by

$$A(t) = \begin{cases} 1 - F(t), & t \in (\tau, \tau + \nu), \\ [1 - F(\tau)]A(t - \tau), & t \in [0, \tau], \\ [1 - F(\tau)]A(t - \tau) + F(\tau)A(t - \tau - d), & t \in [\tau + d, \infty), \end{cases} \quad (13.28)$$

For model A with random renewal durations, the instantaneous availability is expressed as follows:

$$A(t) = [1 - F(\tau)]A(t - \tau) + F(\tau) \int_0^{t-\tau} A(t - \tau - y)g(y)dy \quad (13.29)$$

The authors also give the expression of the stationary availability in this case:

$$UTR = \frac{\tau - \int_0^{\tau} F(x)dx}{\tau + F(\tau) \int_0^{\infty} G(y)dy} \quad (13.30)$$

For model B with constant renewal duration, the instantaneous availability is given by

$$A(t) = F(t) + \sum_{i=1}^{\lfloor (t-d)/\tau \rfloor} A(t - i\tau - d)[F(i\tau) - F((i-1)\tau)] \quad (13.31)$$

For model B with random renewal durations, the instantaneous and the stationary availabilities are expressed as follows:

$$A(t) = [1 - F(\tau)] + \sum_{i=1}^{\lfloor t/\tau \rfloor} [F(i\tau) - F((i-1)\tau)] \int_0^{t-i\tau} A(t - y - i\tau)g(y)dy \quad (13.32)$$

$$UTR = \frac{\int_0^\infty [1 - F(x)]dx}{\tau \sum_{i=1}^\infty i[F(i\tau) - F((i-1)\tau)] + \int_0^\infty [1 - G(y)]dy} \tag{13.33}$$

The authors discuss the properties of these availability functions and compare systematically their results with those obtained by Sarkar and Sarkar (2000). The latter consider in their model, like in the majority of periodic inspection models, that inspections are performed at instants $\tau, 2\tau, 3\tau, \dots$, independently of the duration between inspection and the end of the system renewal action. Cui and Xie (2005) do not work with this assumption; they rather suppose that the inspections are carried out at fixed instants after the end of each renewal action. In fact, they argue that according to the assumption of Sarkar and Sarkar (2000), the system renewal might end just before the successive inspection; in that case the inspection would no longer be useful (see Figure 13.3).

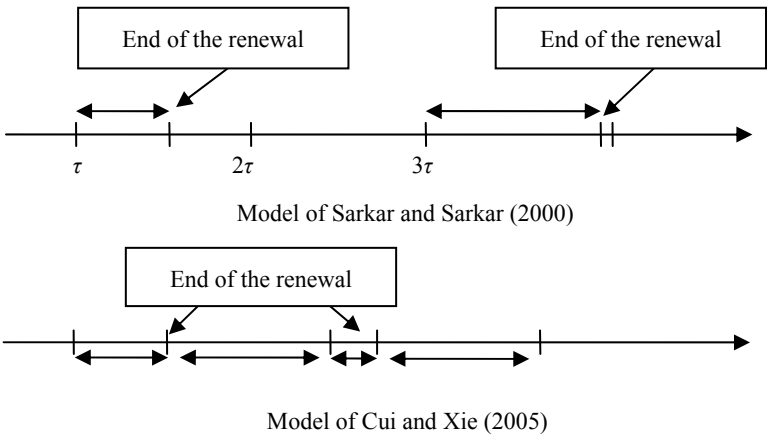


Figure 13.3. Planning of inspection instants according to Sarkar and Sarkar (2000) and Cui and Xie (2005)

13.3.1.6 Systems Alternating Between Periods of Activity and Periods of Inactivity Badia *et al.* (2002) and Chelbi *et al.* (2008) have focused on production systems which alternate between periods of activity and periods of inactivity. They consider situations where failures are instantaneously detected (self-announcing failures) when the equipment is in an operating phase, whereas they can be detected only by inspection during a phase of inactivity (non-self-announcing failures).

The proposed strategy suggests submitting the equipment to inspection when its age reaches T time units. If no failure is detected by inspection, a preventive maintenance is performed. A corrective maintenance action is carried out following failure while the system is in operation, or following the detection of failure through inspection while the system is idle. Inspections may fail and give mistaken results (imperfect inspections). It should be noted that there are two types of imperfect inspection (Gertsbakh, 2000): type I and type II. For type I, which is called ‘error or false positive’, a system is declared in a failed state following an

inspection whereas in fact it is operating. In case of type II, which is called ‘error or false negative’, a system is declared in an operating state whereas in fact it has failed.

Badia *et al.* (2002) establish in these conditions a mathematical model which allows finding the optimal age T at which inspection must be performed minimizing the total expected cost per time unit, whereas Chelbi *et al.* (2008) develop a model which determines the age T which maximizes the equipment stationary availability in the same conditions. In each of these works, the authors establish the conditions of existence and uniqueness of an optimal solution.

The expression of the total expected cost per time unit of Badia *et al.* (2002), and the expression of the system stationary availability, of Chelbi *et al.* (2008) are given as follows:

$$R_c = C_d + \frac{a(T)}{b(T)} \quad (13.34)$$

where

$$a(T) = (C_i + C_p + C_f \alpha)R(T) + [q(C_i + C_p)\delta - pC_p + C_g]F(T) - C_d \int_0^T R(u)du \quad (13.35)$$

$$b(T) = qT[R(T) + \delta F(T)] + (1-q) \int_0^T R(u)du \quad (13.36)$$

The system stationary availability is

$$UTR = \frac{\int_0^T R(u)du}{(1-q) \int_0^T R(u)du + [q(1-\delta)T + C_1]R(T) + q\delta T + C_2} \quad (13.37)$$

where

$$C_1 = (1-q\delta)T_i + (1-q(\delta-1))T_a + \alpha T_f - T_c \quad (13.38)$$

$$C_2 = q\delta T_i + q(\delta-1)T_a + T_c \quad (13.39)$$

13.3.1.7 Other Strategies for Single Component Systems

Hariga (1996) considered periodic inspection of a randomly failing machine. Inspections are supposed to be perfect. Considering on one hand the average profit generated by the use of the machine, and on the other the inspection and repair costs, the author develops the following expression for the average profit per time unit Z as a function of the inspection period τ :

$$Z(\tau) = \frac{p_r \int_0^\tau R(t)dt + C_g R(\tau) - C_i - C_g}{\tau} \quad (13.40)$$

where p_r stands for the average profit per time unit when the machine is operating.

It is shown that there exists a unique period τ which corresponds to zero profit (break-even point) in the case of each type of the failure rate behaviour (IFR, CFR and DFR). The objective is to determine the optimal period τ^* which maximizes the average profit per time unit.

In the same context, Shima and Nakagawa (1984) considered such systems on which protecting devices (components) are installed in order to absorb and avoid internal and external shocks. The system is inspected according to a sequence $X = (x_1, x_2, \dots)$, when the protection device is in a good (operating) state shocks are absorbed with probability $(1-a)$. If the device is in a failed state, then the machine fails following any shock.

The additional following costs are considered:

C_1 : the protective device and the machine failure cost;

C_2 : the protective device cost

with $C_1 > C_2 > C_i$

The optimal inspection sequence X_k is the one which minimizes the total expected cost per time unit $R(X_k; a)$ given by the following expression:

$$R(X_k; a) = \frac{C_1 - (C_1 - C_2) \sum_{j=1}^{\infty} \int_{x_{j-1}}^{x_j} e^{-\alpha a t} e^{-\alpha(x_j - t)} dF(t) + C_i \sum_{j=1}^{\infty} e^{-\alpha x_j} [1 - F(x_j)]}{(1/\alpha) \left\{ 1 - \sum_{j=1}^{\infty} \int_{x_{j-1}}^{x_j} e^{-\alpha a t} e^{-\alpha(x_j - t)} dF(t) \right\} + (1-a) \int_0^{\infty} e^{-\alpha a t} [1 - F(t)] dt} \quad (13.41)$$

Note that the authors use an algorithm developed by Barlow and Proschan (1965) to generate the optimal inspection sequence under certain conditions, particularly when inspection is periodic.

13.4 Inspection Models for Multi-component Systems

In the above-mentioned works, the inspection models are related to systems treated as a single component. A variety of inspection models have been developed in the literature for multi-component systems. Such systems are widely used in industries, for example nuclear and avionic, to name but a few, where systems require very high reliability.

13.4.1 Failure Tree Method Based Strategies

In Reinertsen and Wang (1995) the authors propose an approach for multi-component systems inspection. This approach is developed within a context of failure diagnostics based on a failure tree method. This work is made according to assumptions, namely, (1) each elementary event is perfectly inspected, (2) all elementary events that are binary are independent, and (3) only one event is inspected at a time. A procedure is proposed to derive the optimal inspection sequence. This procedure considers inspections of non-identical durations and short cuts are also of non-identical probability values. Thus, the work of Reinertsen

and Wang generalizes that of Najmus and Ishaque (1994) where shortcuts are considered of identical probability value.

13.4.2 Cases of Cold and Hot Stand-by Systems with Known and Partially Known Lifetime Distributions

In the work of Gopalan and Subramanyam (1982), a standby system composed of two identical repairable components is considered. System component lifetime is exponentially distributed, while time to repair is of a general distribution. Systems as well as the operating component failures are detected only by inspections assumed to be perfect and of negligible durations. The time between successive inspections is a random variable. Initially, only one component is operating while the other is in standby. The system fails if both components have failed. Two types of costs are considered, namely the cost C_1 per unit time when the system is functioning, and the cost C_2 assigned to the system inspection and repair. On the basis of these costs, an analysis of cost vs benefit is conducted. The total benefit is given by the following formula:

$$G_{NET} = \frac{G_T(t)}{t} \quad (13.42)$$

where

$$G_T(t) = C_1 \mu_U(t) - C_2 \mu_B(t) , \quad (13.43)$$

$\mu_U(t)$: is the average time of system functioning in the interval $[0, t]$,

$\mu_B(t)$: is the average time where the system undergoes maintenance.

Different numerical examples have been discussed, considering many combinations of the parameters of the probability distributions associated with times between failures, inter-inspection times and repair times, calculating for each configuration the corresponding net profit.

This model has been extended by the same authors in 1984 (Gopalan and Subramanyam 1984) to consider the case where inspection duration is non-negligible. The model obtained turned out to be much more complex because of the impossibility of obtaining analytical expressions of some inverse transforms of certain functions. The authors resorted to numerical inversion methods.

Cui *et al.* (2004) develop a sequential inspection model for a multi-component standby system. Initially, the lifetime distribution of each component is assumed to be partially known. In this work, system components are simultaneously and sequentially inspected, while a desired system availability level is ensured. At a given inspection, information related to degradation of each component, as well as the estimation of the components lifetime distribution functions, are used to allow the determination of the next inspection time.

Hyo-Seong and Mandyam (1994), consider a standby system composed of N constant failure rate components. Each system component may fail due to random shocks. In this work, the authors present an inspection and a preventive

replacement policy. Roughly speaking, inspections are made on the system components at random times.

At a given inspection time, the decision for the system replacement is made on the basis of the number of failed components, say m , the system is then replaced according to two scenarios:

- The first scenario corresponds to the case where the number m is equal or greater than a given level r ; and
- The second scenario corresponds to the case where all components are failed due to a failure detected without inspection (self-announcing failure).

In this work, replacement times are assumed to be of negligible duration and two types of cost are considered – the cost of system setup induced by the system replacement, and the cost corresponding to the system replacement for a given number m of failed components. It is then shown that there exists a particular number r^* of failed components, the value of which minimizes the average total cost per unit time $TC(r, N)$.

Then the authors propose extensions of their initial model. The latter is modified by considering that replacements are performed at failure. Next, by assuming that the system is put into its down state at each inspection, another model is derived to take into account non-negligible durations and costs of both inspections and replacements. A further extension is proposed where cost induced by system components inventory is considered. Dealing with continuous system supervision and control, Hyo-Seong and Mandyam also propose an inspection policy and develop its corresponding average total cost. The model allows determining wheather it is economically justified to acquire the necessary instrumentation and adopt this policy.

In the same context, Vaurio (1999) showed that for active redundancy systems (hot standby) composed of different components, it is recommended not to inspect all the components simultaneously, but rather stagger inspections at component level. One of the principal reasons is that, when inspections are staggered, the average residence time of a common cause failure is generally shorter than when inspections are simultaneous.

13.4.3 Case of Systems with Components Failure Dependency

The majority of inspection models for multi-component systems assume that components fail independently. As pointed out by Mosleh *et al.* (1998), there are some typical operation conditions where the failure of a given system component induce the total failure of the entire system. Such operation conditions may include physical proximity, similar preventive maintenance procedures, the identical operation principle, and common shared environment, to name a few.

According to this observation made by Mosleh *et al.* (1998), Zéqueira and Bérunguer (2004) study the inspection problem by considering a two components parallel system. System components are characterized by constant failure rates and they are failure dependent, *i.e.*, the failure of a given component may induce the failure of the other component with probability p . The authors derive the system

reliability in the case where inspections are made simultaneously on system components. By varying the value of the probability p , numerical results on test examples are then presented. From these results, as argued by Vaurio (1999), staggered inspections appear more likely to be appropriate than simultaneous inspections of components. This becomes particularly of interest when the objective is twofolds – first, to minimize the average unavailability of the system in a short planning horizon, and second to minimize the inspections cost per time unit.

In order to provide practitioners with supplementary and more direct means to observe the equipment degradation, the following section addresses the concept of conditional maintenance. Such a concept aims to help the maintenance crews in tracking the equipment aging and degradation which may cause the decrease of components performance.

13.5 Conditional Maintenance Models

Periodic preventive maintenance made it possible to reduce considerably the frequency of the accidental breakdowns of the equipment whose operational characteristics deteriorate with age. Such a maintenance policy suggests replacing the equipment in a preventive way after either a predetermined age or at specific moments independently of the equipment age. From an economic point of view, it is of interest to replace the equipment right before the occurrence of the equipment failure. This can be realized, on the one hand, only if the equipment degradation may be tracked, and on the other, if parameters of operational environment are controlled so as to avoid any equipment failure due to such parameters.

In contrast to periodic preventive replacement, conditional maintenance actions are closely related to equipment state. Accordingly, preventive replacement is performed only when an alarm threshold is reached. This makes it possible to reduce the number of the equipment replacements and by the same time to ensure higher equipment availability.

In the next section we present conditional maintenance models for single component systems. We will distinguish models for systems for which the degradation process is assumed to be respectively continuous and discrete.

13.5.1 Conditional Maintenance Models for Single Component Systems

13.5.1.1 Single Component Systems with a Continuous Deterioration Process

Generally, the equipment degradation level is evaluated by performing measurements. In some particular situations, such measurements are experienced on parameters, called control parameters, which are closely related to the process of the equipment degradation. Such control parameters may include for example vibrations magnitude, the degree of acidity of a lubricant or its chromium concentration. A variety of models have been developed in the literature to allow, on the one hand, control parameters identification, and on the other, design and setup of data acquisition and diagnosis systems (Scarf, 1997a, b).

According to the above-mentioned context, several inspection models have been proposed (Hopp and Kyo, 1998, Turco and Parolini, 1984, Pellegrini, 1992 Park, 1988a, b Chelbi, and Ait-Kadi, 1999 Christer and Wang, 1995 Barbera *et al.* 1996 Wang, 2000).

The focus of these models is generally the determination of the optimal sequence of the inspection times for a given alarm threshold, or the optimization of the alarm threshold for predetermined inspection times. Nevertheless, an approach which is widely used consists in modeling the system residual lifetime with respect to the degree of deterioration reached, and then to use this model in an economic model by considering maintenance costs.

Generally, strategies of conditional inspection are of two types, namely type I and type II. These strategies are adopted depending on the environment of the system to be inspected. A strategy of type I is such that the inspection times are determined in advance independently of the result obtained for each inspection. The second type consists of strategies for which the interval between consecutive inspections is updated according to the results obtained by the previous inspections. In the literature, the important advantage assigned to the type II strategy consists in the fact that the number of inspections can be controlled by the determination of the next inspection time, while being based on the result of the current inspection.

However, in the case where several important system components should be inspected and if each inspection would require the stoppage of the system, strategies of type II could induce an important reduction of the system availability. It follows that such strategies may be more appropriate for multi-component systems having only one important component to be inspected (Tsurui and Tanaka, 1992 Toyoda-Makino, 1999).

Toyoda-Makino (1999) considers an inspection strategy for a single component. This strategy allows the detection of possible cracks due to fatigue and whose propagation is random. In this work, the inspection strategy of type II is adopted and at each crack detected the component is assumed to be immediately replaced. The author recalls that in this context of crack random propagation, it is generally difficult to evaluate the inspection effectiveness just after the inspection is performed (Tanaka and Toyoda-Makino, 1998). Nevertheless, the author proposes a method to derive a quantitative evaluation of the effectiveness of inspections. Roughly speaking, at the end of a given inspections sequence $(x_1, \dots, x_n)$, an evaluation is performed at the assessment time T_n . The objective is then to determine an inspection sequence which minimizes the total average cost per unit time $R(s_1, \dots, s_n; T_n)$. The method proposed is based on two steps; the first consists in minimizing the average total cost for a fixed assessment time T_n , while the second determines the optimal value of T_n by minimizing the average total cost per time unit.

By considering the principle of type I inspection strategy, Turco and Parolini (Turco and Parolini, 1984) and Chelbi and Ait-Kadi (1998) have proposed the following inspection policy: the equipment is inspected at times $(x_1, x_2, x_3, \dots, x_n)$ and measurements are performed on one or more important characteristics of the state of the equipment. A replacement by new identical equipment is then carried

out whenever either the characteristics values (measurements) are non-acceptable or the equipment has failed.

The profile of equipment wear is given in Figure 13.4 where three zones are distinguished. Whenever the equipment enters zone III, its failure becomes imminent.

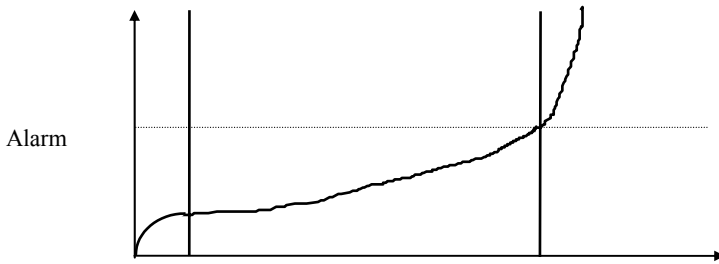


Figure 13.4. Equipment wear profile

The alarm thresholds are generally empirical and equipment sensitive. During an inspection, when the alarm threshold is exceeded, a preventive action is planned. Note that in spite of accurate equipment monitoring, random failure can never be entirely circumvented.

In both works (Turco and Parolini, 1984; Chelbi and Ait-Kadi, 1998), when dealing with wear of equipment, two distinct probability density functions (pdf) $\phi(\cdot)$ and $h(\cdot)$ are assigned to the equipment lifetime. The pdf $\phi(\cdot)$ describes the behaviour of the equipment before the alarm threshold is reached, while the pdf $h(\cdot)$ corresponds to the residual lifetime of the equipment in the case where the alarm threshold is exceeded.

At the i th inspection performed at time x_i , if the alarm threshold is exceeded or the equipment has failed, then the equipment undergoes a preventive or corrective action at time $x_i + H$ (see Figure 13.5). The duration H corresponds to time incurred by administrative procedures and resources preparation for maintenance actions. The durations of maintenance actions are assumed to be negligible.

In the work of Turco and Parolini (1984), the authors propose an inspection model where it is assumed that the equipment state is not affected by the inspection operation. Between consecutive inspections, the conditional probability that the alarm threshold is exceeded is also assumed to be constant. Furthermore, for the sake of simplicity, possible failures are assumed to be instantaneously detected. Chelbi and Ait-Kadi (1998) take into account the fact that inspections may affect the equipment state. They also consider the situation where the equipment state, including failure, can be known only after inspection. This leads to an inspection model which considers an idle period between the time where a failure occurs and that of its detection.

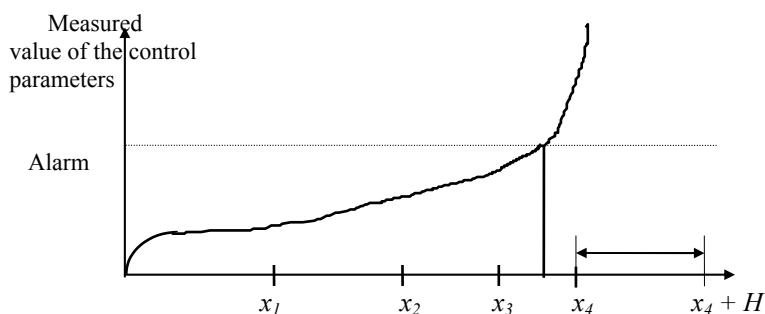


Figure 13.5. Evolution profile of a control parameter with time

Park (1988a) proposes a periodic inspection strategy for a class of systems for which the degradation process is continuous and cumulative with nonnegative, stationary and statistically independent increments. At a given inspection, preventive replacement is performed if the degradation level exceeds a critical threshold r ($0 \leq r \leq b$), b being the value corresponding to the failure that induces the equipment replacement. In this work, inspections and replacements have negligible durations, and the evolution of time is measured in terms of interval between two inspections. Conditions are then derived so as to ensure the existence of an optimal alarm threshold r^* which minimizes the average total cost per unit time, $R_c(r)$, on an infinite horizon.

An example is presented where input parameters are arbitrarily chosen and the degradation process follows a Gamma distribution. By considering different values of the inspection period, curves are given to show the variation of the average total cost per time unit vs the alarm threshold r (Figure 13.6).

In line of the work of Park (1988a), Dieulle *et al.* (2003) consider equipment whose degradation process is continuous and whose time to failure is distributed according to a Gamma distribution. The equipment is inspected according to a given random inspections sequence. Each inspection consists in determining the system state by measuring a given chosen control parameter. Accordingly, a preventive replacement is carried out if the measured value exceeds a threshold M . Whenever the system state reaches a value L ($L > M$), the equipment is considered in its failed state. In this case, a non-planned and costly replacement is required. The authors suppose that the duration between two inspections is a continuous random variable which depends on the system state given by the current inspection. An inspection scheduling function $m(\cdot)$ is then derived. This function has two properties: (1) it is defined from $[0, M]$ to $[m_{min}, m_{max}]$, and (2) it is a decreasing function.

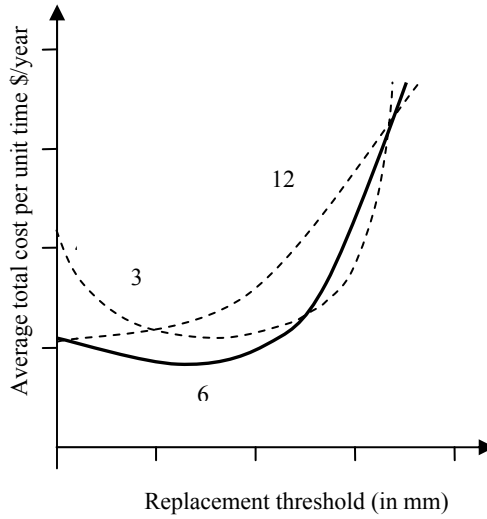


Figure 13.6. Variation of the average total cost per unit time vs the alarm threshold for different inspection periods (Park, 1988a)

The next inspection time x_{n+1} is given according to the following rule (Figure 13.7):

$$x_{n+1} = x_n + m(Y_{x_n}) \quad (13.44)$$

where Y_{x_n} provides the system state immediately after the maintenance action performed at time x_n .

Dealing with conditional maintenance, the majority of the works in general consider as a decision variable either inspection time or the alarm threshold. It is interesting to note that one of the exceptions appears in this work (Dieulle *et al.* 2003) where the authors study the combined effect of the threshold value M and that of the inspection time given by the function $m(\cdot)$. The average total cost per unit time is then function of the two decision variables. To derive such a cost, the authors develop a probabilistic procedure based on the semi-regenerative property of the evolution process. Numerical calculations carried out, show that there is a combination of the two decision variables (M and $m(\cdot)$) which minimizes the average total cost per time unit.

As in Dieulle *et al.* (2003), Wang (2000) developed an approach where the maintenance total cost is function of both the inspection period and the alarm threshold value. Note that the inspection interval is, however, assumed to be deterministic.

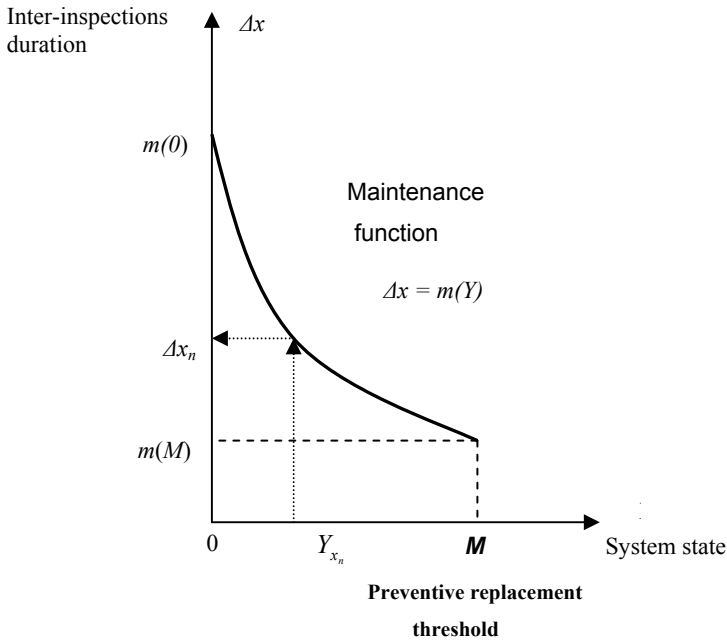


Figure 13.7. Next inspection time determination (model of Dieulle *et al.* 2003)

In Grall *et al.* (2002), Grall and his coauthors propose an inspection model where the degradation process of the system is continuous. By taking the average total cost per unit time as a performance criterion, they optimize simultaneously the inspection time and the threshold level. In this work, the system is assumed to be controlled at instants x_k ($k=1, 2, \dots$), while inspections are considered as perfect and of negligible durations. The system is considered as failed whenever, at a given inspection, its deterioration level exceeds a threshold L . In this case the system is immediately replaced by a new identical one. If the degradation level is found to be higher than a critical threshold, a preventive replacement is then performed.

The next inspection times are chosen with respect to the system state given by the current inspection. At each instant t , the system state is described by a variable $Y(t)$ which varies according to threshold values ξ_k (with $0 < \xi_k < \dots < \xi_N < L$, $\xi_0 = 0$). The value ξ_N corresponds to the critical threshold (Figure 13.8).

Following an inspection performed at time x_k , the procedure is as follows:

- If $\xi_1 \leq Y_k < \xi_{l+1}$ ($0 \leq l < N$), the next inspection time is programmed $N-l$ periods later, *i.e.*, at $x_{k+(N-l)}$, and no other action is undertaken.

- If $Y_k \geq \xi_N$ and $Y_k < L$, a preventive replacement is performed and the system state is reset to the null value. The next inspection is programmed N periods later.
- If a failure is detected, $Y_{k-1} < L \leq Y_k$, then the system is replaced by a new and identical one, and the next inspection is programmed N periods later.

Thus, according to the procedure described above, parameters N and ξ_i ($i=1, \dots, N$) together with the information collected on the system during each inspection are used to determine the next inspection time.

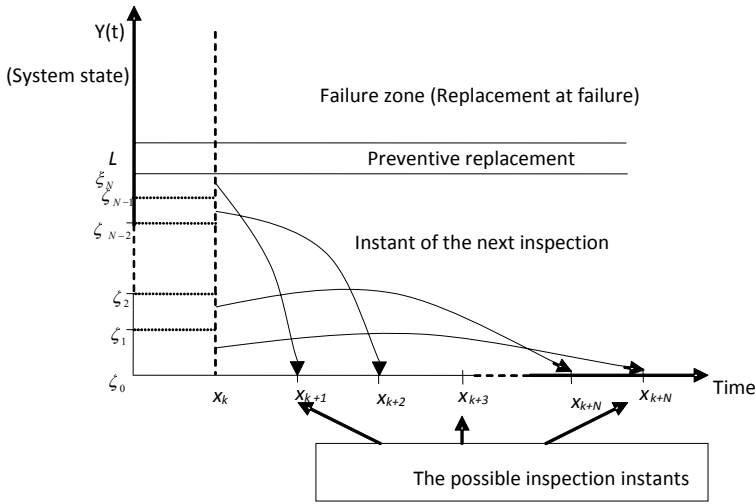


Figure 13.8. Inspection policy of Grall *et al.* (2002)

The average total cost per time unit on a horizon of time Δt is given by

$$\frac{1}{\Delta t} \left(C_i \int_0^{\xi_N} g_i(y) dy + \int_{\xi_N}^L C_p(y) g_i(y) dy + C_c \int_L^{+\infty} g(y) dy + \int_0^{\xi_N} C_o(y) g(y) dy \right) \quad (13.45)$$

where:

$g_i(y)$ is the probability density function associated with the system state Y and having a programmed inspection (in a long run); and

$g(y)$ is the probability density function associated with the system state Y (in a long run).

The authors derive density functions $g(y)$ and $g_i(y)$ with respect to thresholds ξ_k ($0 < \xi_k < \dots < \xi_N < L$, $\xi_0 = 0$).

Through numerical examples, the proposed inspection strategy is compared to classical strategies of inspection and replacement. The results obtained highlight the fact that the proposed strategy can be adapted to several characteristics of a given system, and it is less costly.

The works addressed above in this sub-section deal with systems for which the degradation process is assumed to be continuous. In the literature, several other approaches have been developed to deal with systems for which the degradation process is discrete.

13.5.1.2 Single Component Systems with a Discrete Deterioration Process

For such systems, the inspection models proposed are generally derived on the basis of Markov chains, they attempt to determine the set of states for which the system should be replaced, so that to minimize the average total cost (Lam and Yeh, 1994b; Ohnishi *et al.* 1986; Hontelez *et al.* 1996; Valdez-Flores and Feldman, 1992; Tijms and Duyn Schouten, 1984; Wijnmalen and Hontelez, 1992; Coolen and Dekker, 1995) and (Chen *et al.* 2003).

Lam and Yeh (1994a) proposed a continuous inspection policy where each inspection consists on measuring a given control parameter j . Therefore, the optimal strategy consists in system replacement either when the optimal threshold j^* is reached or the system has failed, *i.e.*, a threshold L is exceeded.

Lam and Yeh (1994b) on the basis of their work in Lam and Yeh (1994a), proposed a periodic inspection policy where inspections, *i.e.*, measurements of the control parameter j , are performed periodically at each τ units of time. At each inspection time $n\tau$, if the system state Y is such that $j \leq Y \leq L$, then the system replacement is carried out. The optimal values τ^* and j^* are then given.

Chiang and Yuan (2000, 2001) proposed continuous and a periodic inspection models. Each model consists in determining optimal threshold levels i^* and j^* ($i^* < j^*$ and $1 < i^* < j^* < L$) such that: if $i^* \leq Y < j^*$, the optimal action to be performed is a minimal repair, while if $i^* \leq Y < L$ the optimal action consists in a replacement or in doing nothing.

Chen *et al.* (2003) consider a system whose degradation states are numbered from 1 to L such that $1 < 2 < 3 < \dots < L$. State 1 indicates the perfect functioning state and L corresponds to the failed state for which the system replacement is required. For each state is assigned a health index H . This index is derived from the system operational characteristics. In this work, the system is periodically inspected each τ units of time. At each inspection, measurements are performed on the system operational characteristics. The system then undergoes a maintenance action W_{ik} which consists in driving it from state i back to the lower state k . An average cost is assigned to a maintenance action.

At the end of each inspection, a real time procedure provides the optimal maintenance action to be performed. This procedure is implemented on a central computer which is connected to the system (see Figure 13.9). At a given inspection, this procedure calculates: the health index H corresponding to the measured values and consequently determines the system state $i \in \{1, \dots, L\}$, and the average cost corresponding to each possible maintenance action W_{ik} . This cost is given on the basis of a transition probability matrix.

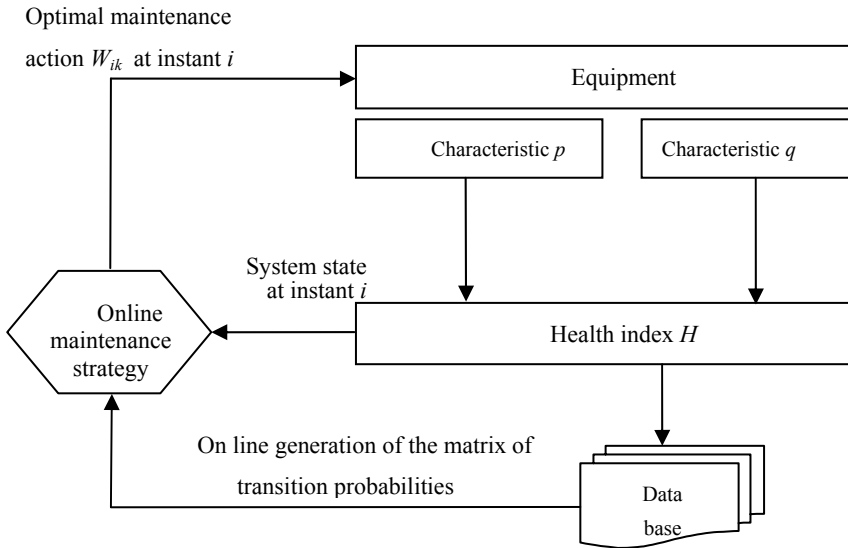


Figure 13.9. Inspection strategy proposed by Chen *et al.* (2003)

13.5.2 Conditional Maintenance Models for Multi-Component Systems

If components of a system are independent, in this case the conditional maintenance problem can be reduced to that of a system composed of a single component. However, as pointed out in Dekker and Smith (1998), Dekker *et al.* (1997) and Wildeman (1996), if system components interact according to either economic, stochastic or structural dependencies, in this case, the optimal strategy for a single system component is not necessarily an optimal strategy for the entire system.

In the literature, the majority of the existing maintenance models for multi-components systems recommend the grouping of the maintenance actions (Dekker *et al.* 1997; Cho and Parlar, 1991; Van der Duyn Schouten, 1996). However, a restricted number of these models are developed within a conditional maintenance context.

In Marseguerra *et al.* (2002), the authors demonstrate the difficulties concerning the extension of a maintenance problem from single-component to multi-component systems. Indeed, analytical modeling becomes difficult and simulations are useful for such situations.

In a multi-component systems setting with economic dependency, two types of conditional maintenance models can be distinguished. The first concerns stationary models which are based on a planning over an infinite horizon, while the second concerns dynamic models where real time decisions are generated. By considering the second type, Wildeman (1996) and Wildeman and Dekker (1997) propose an approach on a rolling horizon. The authors develop a penalty function in order to

compare the maintenance actions grouping method with the classical one, which consists in performing maintenance actions separately on each component.

The model proposed in Castanier *et al.* (2005) extends the model initially introduced in Castanier *et al.* (2001a,b) where a system of a single component is studied. Castanier *et al.* (2005) deal with the problem of conditional maintenance for a series system composed of two components. Each component is subjected to a sequence of periodic inspection. Maintenance actions include preventive replacements or replacement by new identical components. Each maintenance action cost is composed of a fixed cost and a cost specific to the component inspection or replacement. If an action concerns the two components simultaneously, there is a scale economy, since the fixed cost is incurred only once.

A mathematical model is developed to provide a decision-making framework for optimal coordination of the inspection and replacement actions, while minimizing the average total cost per unit time over an infinite horizon. On the basis of the multiple thresholds principle adopted in Grall *et al.* (2002), each component i ($i = 1, 2$) in the model of Castanier *et al.* (2005) is assigned a family of thresholds $\xi_k^{(i)}$ ($0 < \xi_k < \dots < \xi_{n_i} < L_i$, $\xi_0^{(i)} = 0$). For opportunistic replacement purpose, for each component i , additional threshold ξ_i is introduced.

At a given inspection performed at time x_k , the degradation level of each component i is represented by a variable y_i . The maintenance policy is described by a procedure composed of three steps related, respectively, to the component degradation level, the entire system degradation level, and the time of the next inspection. This procedure is described as follows.

Step 1—(component level): the first maintenance action decision is made separately on each component i according to values of its corresponding thresholds family $\xi_k^{(i)}$. Three cases are then possible. The first corresponds to $y_i \in [0, \xi_n^{(i)}]$; in this case no maintenance action is required. The second case is where component i undergoes preventive replacement and corresponds to $y_i \in [\xi_n^{(i)}, L_i)$, while the third case is where a corrective replacement is carried out and corresponds to $y_i \geq L_i$.

Step 2—(system level): according to the new ‘opportunistic replacement’ threshold ξ_i , if the value of the state variable $y_i \geq \xi_i$ and a replacement of component j ($j \neq i$) is programmed, in this case the replacement of component i is simultaneously scheduled.

Step 3—(next inspection time): let y_i^+ be the component degradation level immediately after a maintenance action. The next inspection time of component i is then given with respect to y_i^+ . Thus, if $y_i^+ \in [\xi_k^{(i)}, \xi_{k+1}^{(i)})$, for $k = 0, \dots, n_i - 1$; then the inspection time of component i is planned at $(n_i - k)$ time units later, while in the case where $y_i^+ \in [\xi_k^{(1)}, \xi_{k+1}^{(1)})$ and $y_2^+ \in [\xi_l^{(2)}, \xi_{l+1}^{(2)})$, the next inspection of the entire system is planned at time $t = \text{minimum}\{(n_i - k), (n_2 - l)\}$ decision periods later.

The authors derive on this basis the expression of the total average cost per time unit over an infinite horizon. To check the robustness of their model they present numerical examples by varying values of costs, alarm thresholds, *etc.* The proposed model could be extended for systems composed of more than two components. However, such an extension would lead to a degree of complexity much more important, implying consequently a greater number of decision variables. Therefore, numerical solution of the problem would quickly become very difficult. To overcome this difficulty, it would be interesting to adopt an approach which uses both analytical modeling and simulation.

13.6 Conclusion

This chapter has presented an overview of many contributions dealing with the development of inspection strategies for randomly failing systems. According to the authors' experience, key elements regarding the modelling and the optimization of inspection policies were provided.

The cases of single component and multi-component systems have been considered. Special attention was given to inspection policies in the context of condition based maintenance. The practice of this type of maintenance is particularly interesting because it aims to ensure better monitoring of the equipment degradation process. It consists in providing the practitioners with tools which allow detecting, through inspection, signals of ageing or wear, indicating the imminence of failure. Many technological tools using vibration or lubricating oils analysis, thermography, spectrographic analysis, ferrography, *etc.*, allow one to evaluate the control parameters whose evolution is correlated with the system's state.

However, many issues still have to be addressed by future research on inspection policies. One of the promising avenues consists in dealing with multi-component systems (especially those with more than two components) considering economic, stochastic and structural dependency. Innovative approaches and powerful algorithms are needed to tackle this issue. Regarding conditional inspection policies, further research can be motivated by recent advances in sensor technologies which have not yet been fully utilized to impact dynamic maintenance policies. In fact, low-level degradation data could be used dynamically to effect, in an optimal way, high level maintenance decisions. This could be possible by considering components lifetime distributions that evolve temporally due to the evolution of their degradation mechanisms, which can be measured during inspections.

Finally, it should be noted that all the presented models could be applied in many areas such as medical, military, nuclear and other domains.

References distribution by date and type of model are summarized in Tables 13.1 and 13.2, respectively:

Table 13.1. References distribution by date

<i>By date</i>				
<i>Before 1970</i>	<i>1970–1980</i>	<i>1980–1990</i>	<i>1990–2000</i>	<i>2000–2008</i>
8 references	6 references	12 references	34 references	21 references

Table 13.2. References distribution by type of inspection models and type of systems

<i>By type of model</i>			
<i>Simple inspection models</i>		<i>Conditional inspection models</i>	
Single component systems	Multi- component systems	Single component systems	Multi- component systems
26 references	9 references	33 references	13 references

References

Badia FG, Berrade MD, Campos CA (2002) Optimal inspection and preventive maintenance of units with revealed and unrevealed Failures. *Reliab Eng Syst Saf* 78: 157–163

Barbera F, Schneider H, Kelle P (1996) A condition based maintenance model with exponential failures and fixed inspection intervals. *J Oper Res Soc* 47: 1037–1045

Barlow RE, Proschan F (1965) *Mathematical Theory of Reliability*. John Wiley & Sons, New York: 108–117

Barlow RE, Hunter LC, Proschan F (1960) Optimum checking procedures. *Proceeding of the seventh National symposium on reliability and quality Control*: 485–495

Barlow RE, Hunter LC, Proschan F (1963) Optimum checking procedures. *J Soc Ind & Appl Math* 11: 1078–1095

Castanier B, Grall A, Bérenguer C (2001a) Une structure multi seuil pour l’optimisation de la maintenance des systèmes à détérioration continue. *J Eur Syst Automat* 35: 1235–1262

Castanier B, Bérenguer C, Grall A, Zio E, Demichela M, Piccinini N (2001b) A model for maintenance cost optimization under availability constraints Safety and reliability. *Proc ESREL Toward a safer world 2*: 1313–1320

Castanier B, Bérenguer C, Grall A (2005) A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliab Eng Syst Saf* 87: 109–120

Chelbi A, Ait-Kadi D, Aloui H (2008) Optimal inspection and preventive maintenance policy for systems with self-announcing and non-self-announcing failures. *J Qual Maint Eng* 14: 34–45

Chelbi A, Ait-Kadi D (2000) Generalized inspection strategy for randomly failing systems subjected to random shocks. *Int J Prod Econ* 64: 379–384

Chelbi A, Ait-Kadi D (1999) An optimal inspection strategy for randomly failing equipment. *Reliab Eng Syst Saf* 63: 127–131

Chelbi A, Ait-Kadi D (1998) Inspection and Predictive Maintenance Strategies, *Int J Comput Integ Manufact* 11: 226–231

Chen CT, Chen YW, Yuan J (2003) On a dynamic preventive maintenance policy for a system under inspection. *Reliab Eng Syst Saf* 80: 41–47

- Chiang JH, Yuan J, (2001) Optimal maintenance policy for a Markovian system under periodic inspection. *Reliab Eng Syst Saf* 71: 165–172
- Chiang JH, Yuan J, (2000) Optimal maintenance policy for a production system subject to ageing and shocks. *J Qual Maint Eng* 6: 200–216
- Cho D, Parlar M (1991) A survey of maintenance models for multi-unit systems. *Eur J oper res* 51: 1–23
- Christer AH, Wang W (1995) A simple condition monitoring model for a direct monitoring process. *Eur J Oper Res* 82: 258–69
- Coleman JJ, Abrams IJ (1962) Mathematical model for operational readiness. *J Oper Res* 10: 126–138
- Coolen F, Dekker R (1995) Analysis of a 2-phase model for optimization of condition-monitoring intervals. *IEEE Trans Reliab* 44: 505–511
- Cui LR, Xie M (2005) Availability of a periodically inspected system with random repair or replacement times. *J Stat Plan and Inf* 131: 89–100
- Cui LR, Loh HT, Xie M (2004) Sequential inspection strategy for multiple systems under availability requirement. *Eur J Oper Res* 155: 170–177
- Dekker R, Smith M (1998) New developments in the optimisation of maintenance. *Proc of the European Safety and Reliability Conf ESREL'98, Trondheim, Norway* 731–737
- Dekker R, Frenk H, Wildeman E (1996) How to determine frequencies for multi-component systems ? A general approach *Reliability and maintenance of complex systems, NATO ASI Series. Comput Sys Sc* 154: 239–280
- Dekker R, van der Duyn Schouten F, Wildeman R (1997) A review of multi-component maintenance models with economic dependence. *Math Meth Oper Res* 45: 411–435
- Derman C (1961) On minimax surveillance schedules. *Nav Res Logist* 8: 415–419
- Dieulle L, Bérenguer C, Grall A, Roussignol M (2003) Sequential Condition-Based Maintenance Scheduling for a Deteriorating System. *Eur J Oper Res* 150: 451–461.
- Gertsbakh IB (2000) *Reliability theory with applications to preventive maintenance*. Springer, Berlin.
- Gopalan MN, Subramanyam N (1982) Stochastic behaviour of a two unit repairable system subject to inspection. *MicroelectronI & Reliab* 22: 717–722
- Gopalan MN, Subramanyam N (1984) Stochastic Analysis of a Two-Unit System Subject to Non-Negligible Inspection Time. *Reliab Eng Syst Saf* 7: 133–150
- Grall A, Bérenguer C, Deuille L (2002) A condition based maintenance policy for stochastically deteriorating systems. *Reliab Eng Syst Saf* 76: 167–180
- Hariga MA (1996) A maintenance inspection model for a single machine with general failure distribution. *MicroelectronI & Reliab* 36: 353–358
- Hontelez JAM, Burger HH, Wijnmalen DJD (1996) Optimum condition-based maintenance policies for deteriorating systems with partial Information. *Reliab Eng Syst Saf* 51: 267–274
- Hopp W, Kuo YL (1998) An optimal structured policy for maintenance of partially observable aircraft engine components. *Nav Res Logist* 45: 335–352
- Hyo-Seong L, Mandyam MS, (1994) Optimal Replacement Policies for Systems with Multiple Stand-by Components. *IEEE Trans Reliab* 43: 414–422
- Kaio N, Osaki S (1984) Some remarks on optimum inspection policies. *IEEE Trans Reliab* 33: 227–279
- Kaio N, Osaki S (1989) Comparison of inspection policies. *J Oper Res Soc* 40: 499–503
- Keller JB (1974) Optimum checking schedules for systems subject to random failure. *Manag Sci* 21: 256–260
- Lam C, Yeh R (1994a) Comparison of sequential and continuous inspection strategies for deteriorating systems. *Adv Appl Probab* 26: 423–435
- Lam C, Yeh R (1994b) Optimal maintenance policies for deterioration systems under various maintenance strategies. *IEEE Trans Reliab* 43: 423–430

- Leung FK (2001) Inspection schedules when the lifetime distribution of a single unit system is completely unknown. *Eur J Oper Res* 132: 106–115
- Marseguerra M, Zio E, Podofillini L (2002) Condition-based maintenance optimization by means of genetic algorithms and monte carlo simulation. *Reliab Eng Syst Saf* 77: 151–166
- Mine H, Kawai H (1975) An optimal inspection and replacement policy. *IEEE Trans Reliab* 24: 305–309
- Mosleh A, Rasmuson DM, Marshall FM (1998) Guidelines on modelling common-cause failures in probabilistic risk assessment. NUREG/CR-5485. Idaho National Engineering Laboratory, USA
- Munford AG, Shahani AK (1972) A nearly optimal inspection policy. *Oper Res* 23: 373–379
- Najmus S, Ishaque S (1994) Inspection strategy in fault diagnosis with unequal effort. *Reliab Eng & Syst Saf* 43: 275–280.
- Nakagawa T (2005) *Maintenance Theory of Reliability*. Springer Series in Reliability Engineering
- Nakagawa T, Yasui K (1979) Approximate calculation of inspection policy with Weibull failure times. *IEEE Trans Reliab* 28: 403–404
- Nakagawa T, Yasui K (1980) Approximate calculation of optimal inspection times. *J Oper Res Soc* 31:851–853
- Noonan GG, Fain CG, (1962) Optimum preventive maintenance policies when immediate detection of failure is uncertain. *Oper Res* 10: 407–410
- Ohnishi M, Kawai H, Mine H (1986) An optimal inspection and replacement policy for a deteriorating system, *J Appl ProbAB* 23: 973–88
- Park KS (1988a) Optimal Continuous-Wear Limit Replacement under Periodic Inspections. *IEEE Trans Reliab* 37: 97–102
- Park KS (1988b) Optimal Wear Limit Replacement with wear-dependant failures. *IEEE Trans Reliab* 37: 293–294
- Pellegrini C (1992) Choice of a periodic on-condition maintenance policy. *Int J Prod Res* 30:1153–1173
- Reinertsen R, Wang X (1995) General inspection strategy for fault diagnosis-minimizing the inspection costs. *Reliab Eng Syst Saf* 48: 191–197
- Rodrigues DJ (1983) Influence de la période d'inspection sur les coûts dans l'inspection périodique des systèmes. *Rev Stat Appl* 31: 5–15
- Rodrigues DJ (1990) A new approximation for inspection period of systems with different failure rates. *Eur J Oper Res* 45: 219–223
- Sarkar J, Sarkar S (2000) Availability of a periodically inspected system under perfect repair. *J Stat Plan & Inf* 91: 77–90.
- Savage IR, (1956) Cycling. *Nav Res Logist* 3: 163–175.
- Scarf P (1997a) On the modelling of condition-based maintenance. *Advances in safety and reliability. Proc of the ESREL'97 International Conference on safety and Reliability* 3: 1701–1708
- Scarf P (1997b) On the application of mathematical models in maintenance. *Eur J Oper Res* 99: 493–506
- Shima E, Nakagawa T (1984) Optimum Inspection Policy for a Protective Device. *Reliab Eng Syst Saf* 7: 123–132
- Tadikamalla PR (1979) An inspection policy for the Gamma failure distributions. *J Oper Res Soc* 30: 77–80
- Taguchi G, Elsayed E, Hsiang T (1989) *Quality Engineering in Production Systems*. McGraw-Hill, New York
- Tanaka H, Toyoda-Makino M (1998) Cost-based optimal relation between inspection time and assessment time for random fatigue crack growth. *Prob Eng Mech* 13: 69–76

- Tijms H, van der Duyn Schouten F (1984) A Markov decision algorithm for optimal inspections and revisions in a maintenance system with partial information. *Eur J Oper Res* 21: 245–253
- Toyoda-Makino M (1999) Cost-based optimal history-dependent inspection strategy for random fatigue crack growth. *Prob Eng Mech* 14: 339–347
- Tsurui A, Tanaka T (1992) On history-dependent inspection policy for stochastic fatigue crack propagation. *J Soc of Mat Sci Japan*, 41: 1025–1029 (in Japanese)
- Turco T, Parolini P (1984) A nearly optimal inspection policy for productive equipment, *Int J Prod Res* 22: 515–528
- Valdez-Flores C, Feldman R (1992) An improved policy iteration algorithm for semi-Markov maintenance problem. *IIE Trans* 24: 55–63
- Van der Duyn Schouten F (1996) Maintenance policies for multi-component systems : an overview Reliability and maintenance of complex systems. *Comput & Syst Sci* 154: 118–36
- Vaurio JK (1999) Availability and cost functions for periodically inspected preventively maintained units. *Reliab Eng Syst Saf* 63: 133–140
- Wang W (2000) A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance, *Int J Prod Res* 8: 1425–1436
- Wattanapanom N, Shaw L (1978) Optimal Inspection Schedules for Failure Detection in a Model where Tests Hasten Failures. *J Oper Res* 27: 303–317
- Weiss GH (1962) A problem in equipment maintenance. *Manag Sci* 8: 266–277
- Weiss GH (1963) Optimal periodic inspection programs for randomly failing equipment. *J Res of the National Bureau of Standards* 67B: 223–228
- Wijnmalen D, Hontlez J (1992) Review of a Markov decision algorithm for optimal inspection and revisions in a maintenance system with partial information. *Eur J Oper Res* 62: 96–104
- Wildeman R (1996) The Art of Grouping Maintenance. PhD. Diss. Erasmus University, Rotterdam
- Wildeman R, Dekker R (1997) Dynamic influences in multi-component maintenance. *Int J Qual Reliab Eng* 13: 199–207
- Wortman MA, Klutke GA, Ayhan H (1994) A maintenance strategy for systems subjected to deterioration governed by random shocks. *IEEE Trans Reliab* 43: 439–445
- Yang Y, Klutke GA (2000) Improved inspection schemes for deteriorating equipment. *Probab Eng & Inf Sci* 14: 445–460
- Zéqueira R, Bérunguer C (2004) On the inspection policy of a two-component parallel system with failure interaction. *Reliab Eng Syst Saf* 88: 99–107

System Health Monitoring and Prognostics – A Review of Current Paradigms and Practices¹

Ranganath Kothamasu, Samuel H. Huang, William H. VerDuin

14.1 Maintenance Strategies: Motivations for Health Monitoring

The oldest and most common maintenance and repair strategy is “fix it when it breaks”. The appeal of this approach is that no analysis or planning is required. The problems with this approach include the occurrence of unscheduled downtime at times that may be inconvenient, perhaps preventing accomplishment of committed production schedules. Unscheduled downtime has more serious consequences in applications such as aircraft engines.

These problems provide motivation to perform maintenance and repair before the problem arises. The simplest approach is to perform maintenance and repair at pre-established intervals, defined in terms of elapsed or operating hours. This strategy can provide relatively high equipment reliability, but it tends to do so at excessive cost (higher scheduled downtimes). A further problem with time-based approaches is that failures are assumed to occur at specific intervals.

Figure 14.1 illustrates the typical incidence of failure over the life of equipment. At the left, so-called “infant mortality” failures are plotted. Failure rates are low throughout the useful life of a piece of equipment, and rise towards the end of life.

This curve however doesn’t capture the complex interactions between the components of a system and is loosely based on the assumption that the system progresses (or deteriorates) deterministically through a well defined sequence of states (however, the curve might in some cases be valid even if the sequence is not well defined). This assumption is not true especially in the case of discrete manufacturing systems and other complex environments where seemingly random failure behavior is a function of the changes in the work content, schedule and

¹ Published in The International Journal of Advanced Manufacturing Technology, 28(9): 1012–1024, 2006. Reproduced with permission from Springer.

environment effects, as well as unknowable variations between nominally identical components or systems.

The only way to minimize both maintenance and repair costs and probability of failure is to perform ongoing assessment of machine health and ongoing prediction of future failures based on current health and operating and maintenance history. This is the motivation for prognostics: minimize repair and maintenance costs and associated operational disruptions, while also minimizing risk of unscheduled downtime.

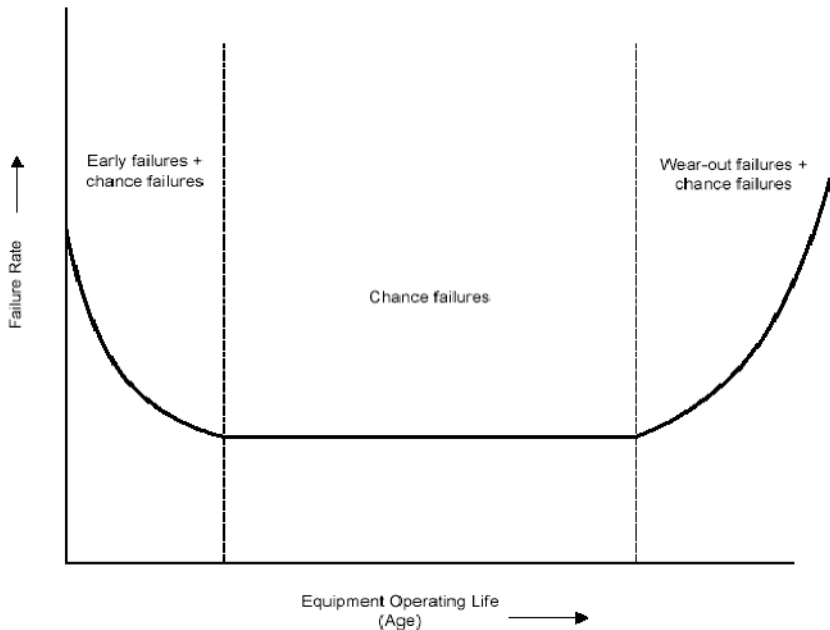


Figure 14.1. Bathtub curve depicting reliability in terms of failure rate of equipment (Stamatis, 1995)

The connection between effective maintenance management techniques and significant improvements in efficiency and profitability has been well documented (Saranga and Knezevic, 2000a). Though the return on investment is highly dependent on the specific industry and the equipment involved, a survey states that an investment in monitoring of between \$10,000 and \$20,000 dollars results in savings of 500,000 dollars a year (Rao BKN, 1996). Across many industries, 15–40% of manufacturing costs are typically attributable to maintenance activities. In the current competitive marketplace, maintenance management and machine health monitoring play an increasingly important role in combating competition by reducing equipment downtime and associated costs and scheduling disruptions (Ben-Daya *et al.* 2000).

Another important motivation for improved maintenance management is its inherent objective to increase machine availability, which has a direct impact on organizational agility. Because of ever increasing customer demands and changes in technology, management strategies such as JIT (Just In Time) and MRP (Material Resource Planning) become essential. These activities improve organizational efficiency by eliminating wasteful production activities. Unscheduled or frequent breakdowns pose a major hindrance to the implementation of such techniques (Abdulnour *et al.* 1995). They also result in high variance in production activities thus increasing the onus on the other business functions such as scheduling. A detailed study of the effects of the maintenance policies on the manufacturing systems is reported in Albino *et al.* (1992), Malik (1979), Reiche (1994), Hans (1999), and Sun (1994).

Another compelling but less addressed justification of maintenance is safety and environmental preservation. With the increase in stringency of safety and environmental laws, proactive maintenance assumes an increasingly important role. Since operational hazards and accidents lead to enormous legal expenses, inattention to these issues is no longer affordable (Rao BKN, 1996).

Quality is increasingly seen as a motivation for improved maintenance management. Since the relation is not immediately apparent it is not surprising that it has not received enough research attention. Since quality improvement is becoming proactive by merging it with techniques like process control and productivity improvement, the effect of equipment maintenance on quality is being exposed (Ben-Daya and Duffuaa, 1995). A similar analysis can be found in Ollila and Malmipuro (1999), Ben-Daya and Rahim (2000), Tapiero (1986), and Makis (1998).

Manufacturing and quality present interesting perspectives to maintenance in the form of objectives to maximize Availability and minimize defective outputs (or in some cases maximize Process Capability). These however are the conjoined objectives of Total Productive Maintenance (Nakajima, 1988), which aims at maximizing Equipment Effectiveness. Terotechnology comes with the same objective but in a much broader sense including the supplier (of the system) and all the involved engineering implementers and users (Husband, 1978).

Because of these insights there has been a considerable shift in perspectives governing maintenance practices in industry. Equally importantly, new theoretical advances and computer-based technologies have provided critical new maintenance management capabilities. These techniques are often interdisciplinary, originating in quite disparate fields. The objective of this chapter is to survey current theories and practices in system health maintenance and to identify relevant references.

Sections 14.2 and 14.3 elaborate on health monitoring paradigms, tools and techniques. Section 14.4 contains a survey of recent case studies that use state-of-the-art techniques in data modeling for machine monitoring, failure prediction and control in industrial applications while Section 14.5 surveys the academic and industrial institutions focusing on the maintenance cause.

14.2 Health Monitoring Paradigms

Health monitoring and its associated functions have been the focus of research and implementation for quite a few years. Through these years they have significantly evolved in terms of governing philosophy, implementation, and enabling advances in technology, modeling techniques and emerging or redefined necessities. The evolution of system health monitoring has an interesting chronological perspective as elaborated by Kinclaid (1987). A brief taxonomy of the various philosophies is given in Figure 14.2.

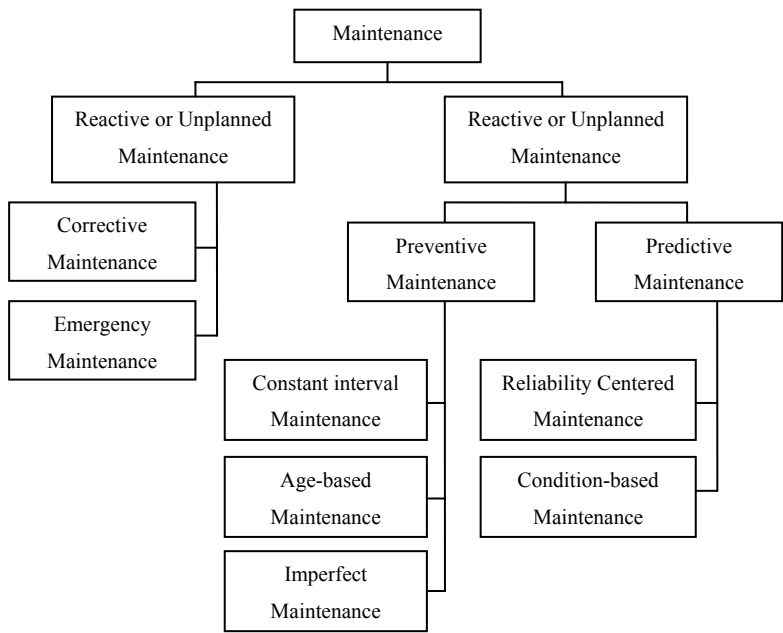


Figure 14.2. Taxonomy of maintenance philosophies

Maintenance philosophies can be broadly classified as reactive and proactive. Reactive or Unplanned maintenance is a legacy practice: maintenance only after the manifestation of the defect, breakdown or stoppage. It is appropriate in facilities where the installed machinery is minimal and the plant is not totally dependent on the reliability of any individual machine (Eisenmann and Eisenmann, 1997). It might also be appropriate when the failure rate is minimal and failure does not result in serious cost setbacks or safety consequence. Breakdown or Corrective maintenance and Emergency maintenance belongs to this category:

1. Corrective maintenance is defined as the activity carried out after a failure has occurred and is intended to restore an item to a state in which it can

perform its required function (Williams *et al.* 1994; Sheu and Krajewski, 1994; Blanchard *et al.* 1995).

2. Emergency maintenance is defined as the maintenance activity that is necessary to accomplish immediately to avoid serious consequences. Constraints are applied on the frequency of maintenance with the object of cost-wise optimization. These constraints are defined in terms of the immediacy of the required action and the possible repercussions of non-maintenance.

Proactive or planned maintenance can be further classified as preventive and predictive maintenance. As the name suggests it does not wait for the equipment to fail before commencing the maintenance operations. In many situations, better utilization of resources are seen compared to reactive strategies (Mobley, 1990).

Preventive maintenance is the strategy organized to perform maintenance at predetermined intervals to reduce the probability of failure or performance degradation. It can be classified into constant interval, age-based or imperfect maintenance:

1. Constant interval maintenance: as the name suggests it is done at fixed intervals (in addition to any maintenance prompted by failure that is performed when it manifests). Intervals are selected to balance high risk of failure with long intervals and high preventive maintenance costs with short intervals (Jardine, 1987).
2. Age-based maintenance: in this strategy, preventive maintenance at fixed intervals is carried out only after the system has reached a specific age, say 't'. If the system fails prior to t, maintenance action is taken and the next maintenance is scheduled to t units later. By deferring initiation, this strategy reduces the number of maintenance intervals compared to constant interval maintenance.
3. Imperfect maintenance: in the above two schemes, the system is assumed to be restored to its original condition after a preventive maintenance. However it may be the case that the condition of the system is in between good (original) and bad (failure). This is the premise of imperfect maintenance strategies which take into consideration the uncertainty of the current state of the equipment while scheduling future activities.

The predetermined interval is estimated from the failure rate distribution that is constructed from historical data extracted from the system or provided by the supplier of individual components in the system. The estimation of distribution and the interval determination are beyond the scope of this paper and the required analysis is extensively covered by Rao SS (1992).

Predictive and preventive maintenance differ in the scheduling of maintenance. In the latter it is performed on a fixed schedule whereas in the former it is adaptively determined. Predictive maintenance can be classified into Condition-based Maintenance and Reliability Centered Maintenance:

1. Condition-based Maintenance (CBM): this is a decision making strategy where the decision to perform maintenance is reached by observing the “condition” of the system and/or its components. The condition of a system is quantified by parameters that are continuously monitored and are system or application specific. For instance, in the case of rotary systems a vibration characteristic or index is an appropriate choice. The advantage of this approach is immediately apparent as the decision is made on depictive and corroborative data that actually reflect the state of the system. It is highly presumptive to assume that the state of a system would always follow the same operational curve, which is the underlying assumption in preventive maintenance. In an industrial or production environment, the system is exposed to random disturbances, which cause deviations in the operational characteristics. Hence it is highly justified to monitor the condition of system and base the maintenance decision on the state of the system. Some of the advantages of CBM are prior warning of impending failure and increased precision in failure prediction. It also aids in diagnostic procedures as it is relatively easy to associate the failure to specific components through the monitored parameters. It can also be linked to adaptive control thus facilitating process optimization. The disadvantage, of course, is the necessity to install and use monitoring equipment and to develop some level of modeling or decision-making strategy.
2. Reliability centered maintenance (RCM): this approach is to utilize reliability estimates of the system to formulate a cost-effective schedule for maintenance. RCM was originally developed in the aircraft industry. For aircraft and other safety-related applications, cost-effectiveness is balanced with safety and availability with the goal of minimizing costs and downtime but eliminating the chance of a failure (Moss, 1985). RCM is a union of two tasks, one of which is to analyze and categorize failure modes based on the effects of the failure on the system and the other is to assess the impact of maintenance schedules on reliability. The failure analysis starts with the identification of all the failure modes and proceeds with categorization of these failure modes based on the consequences of each failure. The results of this study comprise a Failure Modes and Effects Analysis (FMEA). Usually the consequences of failure are Operational, Environmental/Safety or Economic (Rao BKN, 1996). Once the effects have been identified, the decision logic algorithms prioritize the effects. These algorithms tend to be industry specific as the constraints and requirements of each industry vary considerably. Though RCM-based maintenance intervals were determined similarly to planned or scheduled maintenance, condition monitoring techniques are increasingly being used to determine the optimum interval (Kumar and Granholm, 1990; Sandtory, 1991). Hence though originally a preventive maintenance technique, RCM is graduating into predictive maintenance. A good introduction to RCM is given by Moubray (1997), Wireman (1998), Monderres (1993), and Jones (1995).

14.3 Health Monitoring Tools and Techniques

Maintaining the health of a system is a complex task that requires in-depth analysis of the target system, principles involved, and their applicability and implementation strategies. Table 14.1 lists methods, analysis/modeling tools and measurement techniques. However, it has to be noted that most applications are a combination of the listed methods and techniques (tools) and the list is far from exhaustive. For instance, because of their generalized applicability, parameter estimation techniques such as regression, maximum likelihood and expectation maximization can be used in all the listed categories. There is also a close association between reliability-based maintenance and statistical maintenance techniques. A high level explanation of these methods is given in the following sections.

14.3.1 Reliability-based Maintenance

A popular approach to the maintenance of complex systems is through estimating the reliability of the system. Traditionally, reliability is estimating from the time-to-failure distributions of the system. The most striking drawback of such an approach is that multiple failure mechanisms often interact with each other in perhaps unknown ways and this affects the degradation rate of the system, causing it to deviate considerably from the predicted failure distribution. An alternative approach very similar to condition-based maintenance has been proposed by Knezevic (1987) known as the Relevant Condition Parameter (RCP)-based approach. This approach is based on identifying RCPs that quantify or reflect a particular failure mechanism. Using these RCPs the reliability of a system is defined as the probability that RCP lies within prescribed limits:

$$R(t) = P(RCP^{in} < RCP(t_k) < RCP^{lim}) \quad (14.1)$$

RCP^{in} is the initial state of the system and RCP^{lim} is the limiting value where the system inevitably fails. When the failure mechanisms are dependent, it is possible to model the system using Markov chains as shown in Saranga and Knezevic (2000b). Once the Markov chain is formulated representing the different states of the system, the probability of the system being in the upstate $A(t)$ can be calculated as a sequence of integrals of the form given below (Gopalan and Kumar, 1995):

$$A(t) - \int_0^t w(t-x)A(x) = g(t) \quad (14.2)$$

These integrals are further solved by using quadrature techniques such as the trapezoidal approximation.

Table14.1. Maintenance tools and techniques

Methods	Tools	Measurement Techniques
Reliability-based maintenance	Parameter estimation techniques Numerical analysis techniques Markov chains	Vibration analysis Thermography Acoustic emission Wear/debris monitoring
Model-based Failure Detection and Identification (FDI)	State space parameter estimation Artificial neural networks Knowledge-based systems Fuzzy inference systems Neuro-Fuzzy systems	Lubricant analysis Process measurements
Signal-based FDI	Fourier analysis Wavelet analysis Wigner-Ville analysis Diagnostic parameter analysis	
Statistical FDI/ maintenance	Bayesian estimation/ reasoning techniques Markov chains Hidden Markov models Proportional Hazards models	

14.3.2 Model-based Approach to FDI

The model-based approaches to failure detection, isolation and identification (FDI) is based on analytical redundancy or functional redundancy, meaning dissimilar signals are compared and evaluated to identify the existing faults in the system or its components. This comparison is between the measured signal and the estimated values generated by the mathematical model of the system. Figure 14.3 gives a general structure of model-based approaches.

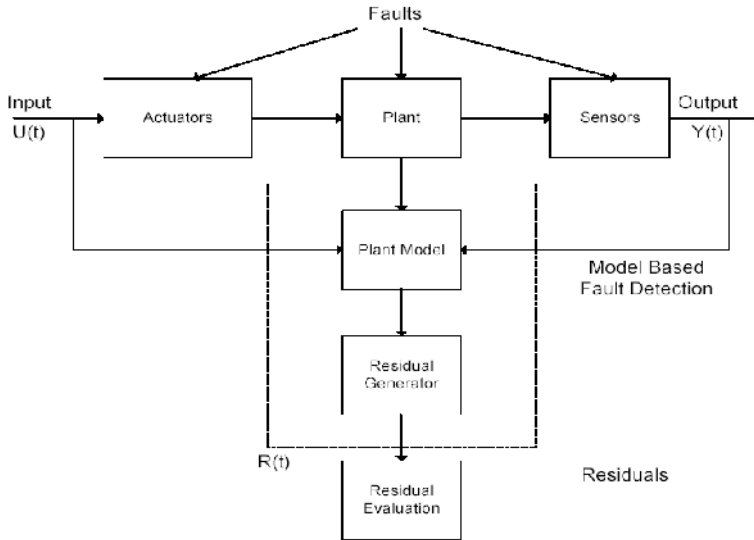


Figure 14.3. General flow of model-based approaches (Simani *et al.* 2003)

Residual generation is the heart of a model-based approach. However, the techniques involved in model-based diagnosis differ in the generation and definition of a residual; for instance in some cases it is the discrepancy of output (from the system) estimation and in some cases it is the error in parameter (of the system's model) estimation itself. It is imperative that the generated residual be dependent only on the faults in the system and not on its operating state. Several techniques that have been proposed in the literature for this residual generation are a modification or improvement of the following three principles:

- Observer-based approaches (Beard, 1971; Ding and Frank, 1990; Patton and Chen, 1997; Wilsky, 1976);
- Parameter estimation technique (Kitamura, 1980; Isermann, 1993); and
- Parity space approach (Chow and Wilsky, 1984; Deckert *et al.* 1977).

Observer-based approaches rely on estimating the outputs from either Luenberger observers or Kalman filters (Simani *et al.* 2003). The approach is centered on the idea that the state estimation error is zero in a fault free environment and it is not so otherwise. Dedicated Observer, Fault Detection Filters and Output Observers are the three important subcategories that fall under this approach.

The basic idea behind the parameter estimation techniques is that the faults affect the outputs through the system parameters. Hence the approach is centered on generating online estimates of the parameters and analyzing the changes in the estimates. In the Equation Error methods which analyze the parameters directly, least square estimation is quite often used; in the Output Error methods which compute the error in the output, numerical optimization techniques are often used.

The principle of Parity Space Relations is to check for parity of the measurements from the process, generating a residual by comparing the model and the process behavior. This approach has been shown to be in close correlation with the observer-based techniques (Patton and Chen, 1994).

As stated before, the model-based FDI approaches are based on identifying (constructing) models that mimic the system. However, an ideal model is never obtained because of nuances of the pragmatic world such as noise, *etc.* and to be effective the model-based FDI should learn to differentiate between these uncertainties and the changes due to failures. Another difficulty is to identify not just the existing faults but the incipient faults which may not (yet) significantly affect the system. Some other useful references are listed (Chen and Patton, 1999, 2000, 2001; 1996b; Patton, 1994; Chen *et al.* 1996a,b; Chow and Willsky, 1984; Ding and Frank, 1991; Duan and Patton, 2001; Nootboom and Leemeijer, 1993; Reiter, 1987; Struss, 1987, 1988, 1989; Struss and Dressler, 1989).

14.3.3 Signal-based FDI

Signal-based FDI approaches focus on detecting the changes or variations in a signal, subsequently diagnosing (identifying) the change. Change detection in a system has been extensively explored in the literature and there are quite a few effective techniques that have integrated various ideas from parametric modeling principles (in statistics) with signal-based principles such as spectral analysis. A good summary is given in Basseville (1988). Some of the techniques are formulated around model-based approaches, *i.e.*, generation of residuals (deviation from nominal signals) and diagnosis of the residuals. Some of the detection algorithms are modeled in the form of hypothesis testing involving a change (or jump) in the mean (known or unknown) such as the Generalized Likelihood ratio test and the Page-Hinkley stopping rule. Some online algorithms are based on computing distance measures between local and global models (differentiated based on their time windows) and some popular measures are the Euclidean distance between AR (Auto Regressive) coefficients, Cepstral distance, chernoff distance, *etc.*

In recent years non-stationary signals have been modeled using wavelets instead of Fourier transforms because wavelets are scale and time variant. Two of the important uses of wavelets to FDI are data compression and feature extraction (Staszewski, 1998). Data Compression as the name suggests refers to encoding the data (like a vibration signal) in a compressed form and feature selection, on the other hand, is identifying features within these encoded signals that would help identify the faults in the monitored systems. Once wavelet transform is applied to the signal output, the coefficients are analyzed for any variation from the normal signal. The identification of coefficients that would substantiate a failure is a painstaking procedure though recently some techniques such as genetic algorithms have been employed. These wavelets are predominantly used for FDI in gears, as vibration analysis is quite effective for these domains (McFadden, 1994; Staszewski and Tomlinson, 1997; Paya *et al.* 1997).

Time-frequency analysis using Wigner-Ville Distribution (WVD) has proven to be another effective tool for vibration analysis. It has proven to be quite effective

in situations where neither the time domain nor frequency domains can produce significant patterns (Staszewski *et al.* 1994). The contour plots generated by WVD are visually inspected for the failure features that indicate its progression and existence. Often these plots are analyzed with the help of classification algorithms ranging from parametric (statistical) to soft computing (neural networks, fuzzy inference systems).

Another popular domain for the application of signal-based FDI is bearing condition diagnosis, where the signals vary immensely because of variable load. These applications typically require signal enhancement via filtering followed by a condition parameter monitoring (Shiroshi *et al.* 1997). Shiroshi *et al.* report an interesting study on the effectiveness of various condition parameters such as kurtosis, crest factor and the peak ratio. Some useful references are listed (Logan and Joseph, 1994; McFadden and Smith, 1985; McFadden, 1986, 1987; McFadden and Wang, 1993, 1996; Eshleman, 1999; Geng and Qu, 1994; Lin and Qu, 2000; Wang, 2001).

Detection signal techniques are also used for FDI, where a detection signal is used as an input to the system for a specific period of time and the diagnosis is based on the behavior of the system during this period. Some interesting theories in the design and implementation of detection signals are given in Nikoukhah *et al.* (2000), Zhang (1989), Kerestecioglu (1993), Kerestecioglu and Zarrop (1994), Uosaki *et al.* (1984), and Nikoukhah (1998).

14.3.4 Statistical FDI/Maintenance

A vast number of applications also use Bayesian statistics and Bayesian parameter estimation for FDI. Some interesting algorithms are proposed by Berec (1998), Won and Modarres (1998), Wu *et al.* (2001), Leung and Ramanougli (2000), and Ray *et al.* (2001). Another important aspect is to identify the detection (inspection) intervals, optimization of cost and replacement decision-making. Markov chains seem to be increasingly used for optimizing maintenance strategies and some algorithms are given in Wang and Sheung (2003), Al-Hassan *et al.* (2000, 2002), and Zhang and Zhao (1999). Another interesting application is given by Bunks *et al.* (2000) using hidden Markov models.

Proportional Hazards Modeling (PHM) has also been used for reliability estimation and estimation of effects on failure rate ever since they were used by Feigl and Zelen (1965). Some interesting theories and applications related using PHM are reported in Jardine (1987), Kobbacy *et al.* (1997), and Pena and Hollander (1995).

14.4 Case Studies in System Monitoring and Control

In a production environment, failure detection is often achieved by operators continuously monitoring the operation of their system and by observing sensor data generated by that system. They recognize significant trends in the operational states of the system and maintain the system by associating trend patterns with failure modes. Challenges with this approach include the loss of knowledge when

the operator changes employment or retires. Often, the difference in levels of expertise induces considerable operational variability by individual operators which is detrimental to the system stability. A second challenge may be a large volume of sensor and operational data resulting in a cognitive overload for the operator, increasing the likelihood of the operator reaching an incorrect decision. A third challenge is that direct observation of operations and available sensor data may not be sufficient to detect incipient failure. Advanced modeling tools are particularly beneficial in these cases, enabling detection and prediction of failures by discovering ways that, for example, the simultaneous variation of multiple parameters is an indication of future failure while the observation of parameters individually (by a person or a computer-based system) cannot provide that insight.

In this section we describe applications of system monitoring in a variety of industrial applications. Adaptive machine and process health models, as employed for predictive monitoring, also enable determination of appropriate changes in control actions to accommodate changes in machine health and environmental conditions. For this reason, we also review applications in adaptive control in this section.

Elements common to all of these applications are the acquisition of data, but more importantly the acquisition, organization and reuse of knowledge to interpret that data and other observables to enable prediction of failure or adjustment of control actions. Differences in the applications include the knowledge acquisition approaches, with common approaches including gathering of rules from domain experts (creating a rule-based system) and discovery of relationships in data (employing neural nets, fuzzy logic and other soft computing approaches).

Villanueva and Lamba (1997) applied the Knowledge-based (rule-based) System (KBS) approach to failure diagnosis in coal processing equipment. They employ the principles propounded by the KADS methodology (Hickman *et al.* 1989) to implement their KBS. The knowledge module of their control model (AshMod) comprises of two components: (1) Goal tree – Success tree (GTST) and (2) Fault-Cause network. The GTST knowledge encapsulates the plant's intentional aspects through a problem reduction strategy and is modeled by a tree structure of hierarchically related goals and sub-goals that must be satisfied for the correct operation of the plant. The upper section (GT) consists of goals and sub-goals that capture purpose. The lower section (ST) consists of success criteria that capture functionality. The fault-cause network is a directed acyclic graph of production rules. This network provides the means for identifying the root cause of an observed fault in the plant. Trend analysis on the GTST tree identifies an abnormal trend and then AshMod uses a mixture of backward and forward reasoning to identify the causes from the fault cause network.

Missed alarms and false alarms are significant problems in diagnostic and prognostic systems, because they diminish the credibility of the monitoring system and diminish the motivation to use or maintain the monitoring system. Although KBS technology can predict failure modes with reasonable accuracy it is not free from these drawbacks. Alonso *et al.* (1998) recommend intensive monitoring of the state along with KBS for more precise diagnosis and prediction. AEROLID, used to monitor a beet-sugar manufacturing plant, compares monitored variables to fixed thresholds, or constant trajectories. The monitoring module thus operates in a

stationary, not adaptive, mode for a given production level range. The system diagnosis may be unstable in the situation of crisp (not fuzzy) thresholds due to diagnoses that vary when encountering small changes in the value of the variable around the threshold. To avoid this instability the authors (Alonso *et al.* 2001) implemented three thresholds for variables, trigger, confirmation and recovery thresholds, which coupled with the temporal information governs the value of the attribute state of the monitored variable. As soon as the variable crosses the trigger threshold, the state changes from OK to vigilance; if it remains over the confirmation threshold for a certain amount of time it becomes critical. When it descends below the confirmation threshold, it changes to vigilance, and when it stays for a given time under the recovery threshold, the state is OK again. The transitions towards vigilance are determined by the threshold. The transition from vigilance to critical always includes a persistence condition over a threshold different from the triggering threshold, which is usually enough to avoid unstable monitoring. If the system state is OK, parameters are only checked against trigger threshold (normal monitoring). When the state jumps into a vigilance model, the intensive monitoring mode is invoked at a higher frequency.

An interesting knowledge representation has been implemented by researchers modeling the proactive maintenance tasks in nuclear power plants. Arroyo-Figueroa *et al.* (2000) used a Temporal Nodes Bayesian Network for failure diagnosis. Each node in the network represents a state attribute along with the temporal information and the edges of the network determine the causal and temporal dependencies among the state attributes. This network is used for real-time fault diagnosis.

Balle and Fuessel (2000) performed Fault Diagnosis and Detection (FDD) by identifying three potential symptom types that can be used in FDD. Assuming the process has already been modeled, a Residual-based symptom is simply the output error between the model and process within a time window of appropriate length. Signal-based symptoms are derived by defining different control performance indices (CPI). Model-based symptoms are derived from the parameters defining the process behavior. Once these symptoms have been quantified, they support a fault isolation scheme by using fuzzy classification trees. The premises in the rules are also controlled by using a process they term as “rule growing”.

A crucial factor for the good functioning of a heuristic diagnosis system is the precision in the knowledge it employs. The design process of the knowledge base is associated with many problems such as inconsistency of knowledge, contradictions between knowledge rules, structural relations between knowledge chunks, causality problems, and representation of knowledge (Slany and Vascak, 1996). Thus development of consistency checkers becomes crucial for the design and continuous improvement of the knowledge base. In building the knowledge base the expert should recognize all possible situations and predict the plausible consequences and describe all the relations among single production rules. This forms a structural hierarchical tree where the roots represent inputs and leaves represent the outputs. In the early stages of knowledge extraction only the rough structure is created and the parameters are only approximated. During the later stages when the expert lays out new rules the compatibility with the old structure should be quantified to check analytically for consistency and contradictions. The

authors (Slany and Vascak, 1996) propose a comparative operator, which makes autonomous consistency checking highly plausible.

Vibration monitoring is widely used for failure prediction. One large class of its applications is rotary machines or components such as gearboxes and bearings. The vibration signal acquired from an accelerometer is able to distinguish between that of a machine or a structural component in good condition *vs* one in which degradation of rolling elements has begun (Paya *et al.* 1995). Many of the spectral analysis techniques such as windowed Fourier transforms (WFT), power spectrum analysis, Wigner-Ville distribution and wavelet transforms have been used in developing fault diagnosis techniques. The features from the spectral analysis are fed into a neural network to classify the observed as fault trends. This approach is favorable as it carries the advantage of being able to classify more than one fault trend in the signal.

Yam *et al.* (2001) employ a similar approach with the exception of using recurrent networks instead of feed forward network for diagnosis. In their case study conducted in utility company, they also developed a maintenance advisory chart which relates the predicted condition of the system with the type of maintenance activity that has to be performed.

An interesting modification in the structure of neural networks used for failure diagnosis was also employed by Gideon (1998). In this structure the input layer represents the potential faults of the system and first hidden layer represents the possible machine components that might be responsible for the possible failure of the system. The second hidden layer stores the operational ranges of these machine components and finally the third hidden layer represents the current operational values of these components. Whenever the current operational value of a specific component is beyond the bounds of its pre-defined approved operational range, it is diagnosed to be the reason for the machine's malfunction.

Leger *et al.* (1998) investigated a fusion between statistical control charts and neural networks for the purpose of FDI. The author also compares the efficiency between multi-layer perceptrons and radial basis networks. It was found that, though the radial basis network requires an inordinate number of hidden layers for the purpose of estimating the failure trends, its efficiency is much better with fewer false alarms.

Process control or optimization is a related task to that of proactive maintenance. Usually the optimized states of a machine are unequivocally defined, but the association of any variation from the optimized state to the multitude of factors that influence the system is a difficult task. The difficulty arises from the complex relationships among the factors themselves and their varied influences on the system. The interactions in general are highly non-linear and pose a tremendous challenge to the modeling technique. Conventional PID (Proportional Integral Derivative) control algorithms may be used to deal with parametric control situations but this approach is difficult to implement.

As an alternative control strategy, Lau *et al.* (2001) effectively employed a combination of neural networks and fuzzy inference systems to control a typical heat transfer system in its optimized state. The system studied consisted of a duct into which heat is conducted by six ribs. The control parameters are spacing between the ribs and the width of the ribs. Neural networks were used first to learn

functional relationships between the desired and controlled parameters. A fuzzy inference system (FIS) with heuristics that can identify the optimum parameters from the predicted control parameters is constructed. The optimization is achieved by activating the neural network with the current operation state of the system thus eliciting the required values of the control parameters for possible optimization. Then the parameter values are refined by using the FIS to predict the optimum change required in terms of the percentage of changes.

Hussain (1999) categorizes the nonlinear control strategies into predictive, inverse model-based and adaptive control and reviews the utilization of neural networks in these categories. According to the author non-linear predictive control refers to the situation where the system, performance objective and the constraints are non-linear functions of the system variables. The inverse model is further divided into direct inverse control and internal model control techniques. In the case of direct inverse control the neural network model has to learn to provide the desired control parameters for the desired targets. Alternatively, in inverse model control the control signals are computed by inverting the forward network model through Newton's method or substitution methods-based on the contraction-mapping algorithm. Adaptive control is further classified into direct and indirect adaptive control. The author lists many successful ventures in each category.

Though these algorithms have been shown to work effectively, they do not effectively utilize the heuristics employed by the operators in process optimization. Their highly unstructured learning tends to produce unstable models. CONES (Connectionist Expert System) is a programming environment designed to capture the expertise of an individual operator for his/her experience on a specific machine (Almutawa and Moon, 1999). The connectionist networks are trained using back propagation principles by following the on-line corrective control actions taken by the experienced operator. Following training, connectionist representation is integrated with a rule-based expert system representation to model the process while an incremental learning technique is used to train the networks further. The network outputs and the weights between the processing elements are fed into the expert system for use in conflict resolution technique (Weight-based Conflict Resolution) that determines the control signals.

The difficulty in process optimization stems from the stringent and time varying requirements for the smooth functioning of a system. Current normal operating standards might result in a hazardous state in the future (Enbo *et al.* 1998). Though fuzzy inference systems have been conventionally used in process monitoring, they tend to assume that the real time variations in the requirements are embedded in the employed heuristics. The authors present an approach where the fuzzy inference system is more dynamic in nature by using time varying membership functions that describe the outputs or the optimized process parameters. This methodology has been successfully adopted in a chemical pulp mill.

Process monitoring is a typical function approximation problem and hence fuzzy models can be effectively employed. Takagi-Sugeno models are a type of FIS where the consequents (outputs) of each rule are usually a linear combination of the input parameters. A drawback of this approach is the increased number of model variables to be identified. Ying (1998) proposes an approach to combat this

problem. In his approach the consequents of only one rule are a linear combination of the inputs. The consequents of the rest of the rules are assumed to directly proportional to the fuzzy output from the first rule related by proportionality constant. This results in a considerable reduction of the model parameters as shown by the author.

Problems such as the knowledge acquisition bottleneck and fuzzy parametric tuning were explored by researchers in Gorzalczany (1999) and Kazamian (2001).

14.5 Organizations and Standards

In recent years there have been noteworthy academic and industrial contributions to the design and implementation of maintenance applications. These advancements are propelled by diverse needs and hence the techniques and algorithms have been fine-tuned to meet specific challenges. Tables 14.2 and 14.3 give a summary of various academic research centers and industries that are targeting the maintenance arena.

Apart from individual efforts, some organizations have come together to streamline the current and future developments in the maintenance and control arena. These alliances have created specific goals and visions to enable the creation of smarter yet flexible tools.

Table 14.2. Academic research centers focusing on maintenance

Research center	Focus	Institution(s)
Center for Intelligent Maintenance Systems (IMS)	Intelligent maintenance, smart sensors, remote and web enabled maintenance	University of Wisconsin Milwaukee; University of Michigan, Ann Arbor
Applied Research Laboratory	Advanced sensing, diagnostics, prognostics and modeling	Pennsylvania State University–State College, Pennsylvania
Condition-based Maintenance Laboratory	Mathematical modeling, statistical analysis, software for CBM applications	University of Toronto, Toronto, Canada

Table 14.3. Survey of industries focusing on maintenance

Company	Capability	Software	Application	Modeling technique
Smart Signal www.smartsignal.com	Sensor data analysis Empirical modeling Offline/Online	eCM 2.0 (run-time engine, database, watchlist manager)	Aviation, power plant, commercial transportation	34 patented algorithms (statistical and time-series)
Data Systems and Solutions www.ds-s.com	Specialized oil analysis systems (JetScan) Mostly offline	CETADS for engine diagnosis, Asset Management software, CAFTA for fault tree analysis	Aviation, hydraulic systems, continuous production systems	Neural networks
Science Applications International www.saic.com	Inspection automation Ultrasonic analysis	IDEAS (data analysis), C-SCAN (inspection automation)	Aviation, industrial safety and security	Neural networks
IVARA www.ivara.com	Reliability centered maintenance	IVARA.ERS (expert system)	Unknown	Expert Systems
Aladon www.aladon.co.uk	Reliability centered maintenance solutions	RCM toolkit (report generation)	Unknown	Statistical
EMA INC www.ema-inc.com	Control system engineering CMMS solutions Information technology	SCADA (supervisory control and data acquisition)	OEM, food processing, Medical, communications	Statistical, Alarm generation
FLIR Systems, INC www.flir.com	Predictive maintenance Infrared cameras for maintenance Automation equipment OEM applications and research	Infrared data analysis and reporting software	Paper, food, steel, petrochemical industries	No modeling

Table 14.3. (continued)

AssetPoint www.tabware.com	Data acquisition and open systems storage Enterprise Asset Management	Tabware (CMMS application)	Petrochemical, OEM Continuous production systems	No modeling
BE&k www.bek.com	Maintenance solutions	Unknown	Paper and pulp Mining, telecommunicati ons power, chemical and processing	Statistical
DRM Technologies www.drmtech.com	Lean manufacturing maintenance solutions	No Software	UNKNOWN	No modeling
General Physics Corporation www.gpworldwide. com	Predictive maintenance Equipment reliability improvement	No Software	Aviation, paper & pulp, telecommunicati ons, high-tech, food & beverage	Statistical
HSB Reliability technologies www.hsbrt.com	World class reliability Total plant reliability Preventable maintenance CMMS solutions Root Cause Failure Analysis	No Software (partners with IVARA)	Unknown	Unknown
Reliasoft www.reliasoft.com	Reliability Assesment FMEA	Weibull++, Blocksim, Xfmea, RG		Weibull analysis Simulation

One such development known as OSA-CBM (Open System Alliance CBM) is being advocated in an effort to integrate various maintenance efforts and to broaden the scope of maintenance by achieving a seamless integration with the different functional arms of a production facility. The initial efforts of OSA-CBM were aimed at developing open and interchangeable systems that cater for the

maintenance arena (specifically CBM systems). The objective is to specify a standard on these various systems so that the future developments tend towards multi-purpose, swappable components. On the software front, COM (Component Object Model) and DCOM (Distributed Component Object Model), CORBA (Common Object Request Broker Architecture), XML (Extensible Markup Language) are being propagated as plausible candidates.

OSA-CBM has developed a seven layered architecture that encompasses the typical stages in the development, deployment and integration of maintenance solutions under the CBM framework. The architecture is depicted in Figure 14.4.

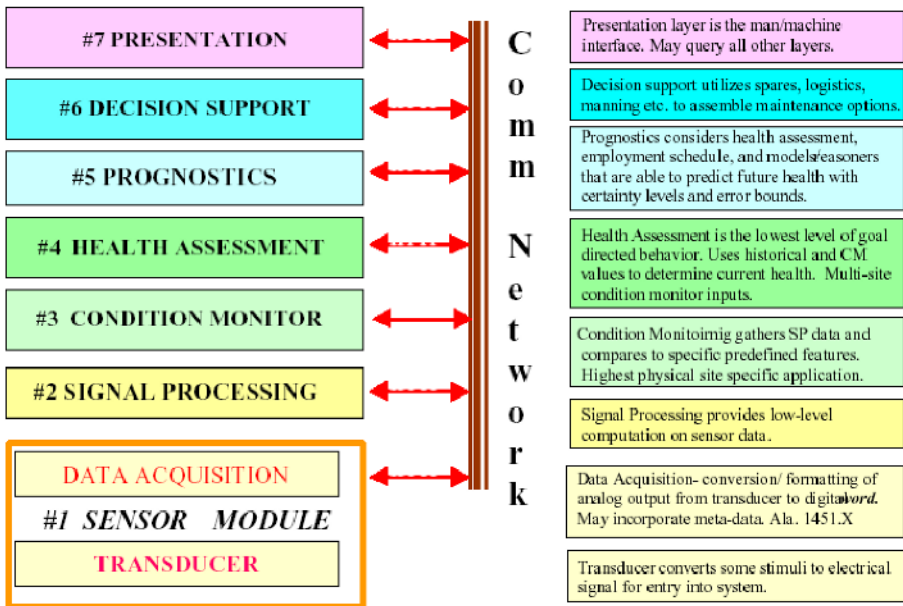


Figure 14.4. OSA-CBM architecture (Lebold *et al.* 2003)

MIMOSA (Machinery Information Management Open Systems Alliance) is an alliance that enables Enterprise Asset Optimization (EAO) resulting from the integration of building, plant and equipment data into and with Enterprise Business Information. OSA-CBM and MIMOSA orient their research and activities to integrate individual goals. The primary objective of MIMOSA is to project maintenance management as a business function that operates on business objects with well defined properties, methods and information interfaces. To this end, it has developed Common Relational Information Schema (CRIS) that specify the equipment database schema as well as the SQL query interface through which data is transferred. This data is typically extracted from condition assessment, control, maintenance management and enterprise information system modules. Figure 14.5 depicts the MIMOSA view of organization elements relevant to the establishment of a maintenance management system. To support the goal of MIMOSA to

integrate these elements in the production environment, MIMOSA interfaces among these elements support four modes of data transfer : file, bulk, SQL (Structured Query Language) and Object.

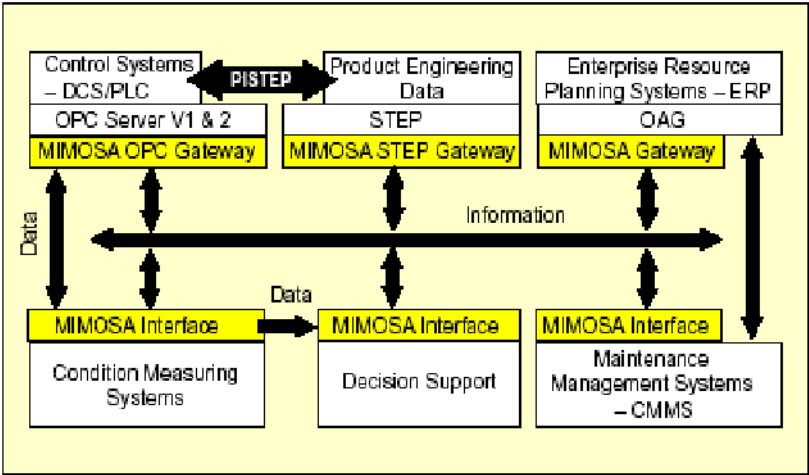


Figure 14.5. MIMOSA and its interfaces (Mitchell, 1998)

14.6 Summary and Research Directions

Maintenance of a system usually starts with the objective to minimize the catastrophic failures that cripple the system. Though time-based maintenance or breakdown maintenance are simpler to implement, condition-based maintenance is gaining popularity because of its proactive approach. Condition-based maintenance is a detailed analytical process that requires in some cases elaborate instrumentation and in most cases complicated modeling techniques. So, it is quite necessary to carry out a requirement analysis prior to implementation of such an effort.

Though the challenges in process control and fault diagnosis are different, artificial intelligence approaches have been shown to be effective for both applications. The modeling of the failure mechanism or process control starts with data collection. Data cleansing is extremely important, particularly in the case of adaptive control. Machine models are usually developed for both applications, and are validated and improved to maintain accuracy and reduce incidence of false alarms and missed hits.

Issues in the development and maintenance of prognostic systems include the selection of knowledge acquisition and modeling technologies, with considerations including available types of knowledge and approaches to achieve and maintain accuracy of the models and knowledge bases. One of the least concentrated efforts in the maintenance arena has been to create applications that are user friendly. These applications tend to be so complex (both in volume and substance) that they

easily overwhelm the user. Given such a scenario, it is not surprising that the user develops deep rooted mistrust in the monitoring system whenever it results in a false alarm or missed hit. Hence, research opportunities include development of modeling technologies that are precise, adaptive, comprehensible and configurable (by user). There is also an opportunity to integrate the qualitative information that can be extracted from FMEA (Failure Mode and Effects Analysis) or FTA (Fault Tree Analysis) of a process or machine into the quantitative analysis that generates diagnostic recommendations.

Acknowledgements

This material is based upon work supported by the National Institute of Standards and technology under Contract No.SB1341-02-C-0049 to VerTech LLC for developing an Intelligent Condition-based Maintenance System. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institute of Standards and Technology.

References

- Abdulnour G, Dudek R A, Smith M L (1995) Effect of Maintenance Policies on the Just-in-Time production System. *Int J of Prod Res*, 33: 565–583.
- Albino V, Carella G, Okogbaa OG (1992) Maintenance policies in Just-In-Time Manufacturing Lines. *Int J of Prod Res*, 30: 369–382.
- Al-Hassan,K, Swailes DC, Chan JFL, Metcalfe AV (2000) Markov models for promoting total productive maintenance. *Proc Ind Sta In Action*, 1: 1–12.
- Al-Hassan K, Swailes DC, Chan JFL, Metcalfe AV (2002) Supporting maintenance strategies using markov models. *IMA J of Manag Math* 13: 17–27.
- Almutawa S, Moon YB (1999) The Development of a connectionist expert system for compensation of color deviation in offset lithographic printing. *Artif Intell in Eng* 13: 427–434.
- Alonso GC, Acosta G, Mira J, De Prada C (1998) Knowledge-based process control supervision and diagnosis: the AEROLID approach. *Expert Syst with Appl* 14: 371–383.
- Alonso GC, Pulido JB, Acosta LG, Llamas BC (2001) On-line industrial supervision and diagnosis, knowledge level description and experimental results. *Expert Syst with Appl* 20: 117–132.
- Arroyo-Figueroa G, Alvarez Y, Sucar E (2000) SEDRET – an intelligent system for the diagnosis and prediction of events in power plants. *Expert Syst with Appl* 18: 75–86.
- Balle P, Fuessel D (2000) Closed –loop fault diagnosis-based on a nonlinear process model and automatic fuzzy rule generator. *Eng Appl of Artif Intell* 13: 695–704.
- Basseville M (1988) Detecting changes in signals an systems – A Survey. *Int Fed of Autom Control*, 24(3): 309–326.
- Beard RV (1971) Failure accommodation in linear systems through self-reorganization. PhD Thesis, MIT, USA.
- Ben-Daya M, Duffua, SO (1995) Maintenance and quality: the missing link. *J of Qual in Maint Eng* 1(1): 20–26.

- Ben-Daya M, Rahim MA (2000) Effect of maintenance on the economic design of x- chart. *The Eur J of Oper Res* 120: 131–143.
- Ben-Daya M, Duffua SO, Raouf A (2000) Overview of Maintenance Modeling Areas. Maintenance, Modeling And Optimization, Kluwer Academic Publishers, Massachusetts, pp 3–35.
- Berec L (1998) A multi-model method for failure diagnosis and detection: Bayesian solution, An introductory treatise. *Int J Adapt Control Signal Process*, 12: 81–92.
- Blanchard B, Verma D, Peterson EL (1995) Maintainability: A key to effective serviceability and Maintenance Management. John Wiley & Sons, New York, NY.
- Bunks C, McCarthy D, Al-Ani T (2000) Condition-based Maintenance of machines using hidden markov models. *Mechanical Systems and Signal Processing*, 14(4): 597–612.
- Chen J, Patton RJ (1999) Robust Model-based Fault Diagnosis for Dynamic Systems. Kluwer Academic, New York.
- Chen J, Patton RJ (2000) Standard H filter formulation of robust fault detection. 4th IFAC symposium on Fault Detection, Supervision and Safety for Technical Processes, 1: 256–261.
- Chen J, Patton RJ (2001) Fault-tolerant control systems design using the linear matrix inequality method. *European Control Conference, ECC'01:1993–1998*.
- Chen S, Billin AS, Cowan CFN, Grant P (1990) Practical identification of NARMAX model using radial basis functions. *Int J Control* 52: 1327–1350.
- Chen J, Patton RJ, Liu GP (1996a) Optimal residual design for fault-diagnosis using multi-objective optimization and genetic algorithms. *Int J Syst Sci*, 27(6): 567–576.
- Chen J, Patton RJ, Zhang HY (1996b) Design of unknown input observer and robust fault detection filters. *Int J Control*, 63(1): 85–105.
- Chow EY, Wilsky AS (1984) Analytical redundancy and the design of robust failure detection systems. *IEEE Trans on Autom control* 29(7): 603–614.
- Deckert JC, Desau MN, Deyst JJ, Wilsky AS (1977) DFBW sensor failure identification using analytic redundancy. *IEEE Trans Autom Control* 22: 795–809.
- Ding X, Frank PM (1990) Fault detection via factorization approach. *Syst Control Lett* 14(5): 431–436.
- Ding X, Frank PM (1991) Frequency domain approach and threshold selector for robust model-based fault detection and isolation. In *Preprint of IFAC/IMACS symposium SAFEPROCESS'91*, 1: 307–312.
- Duan GR, Patton RJ (2001) Robust fault detection using Luenberger-type unknown input observers: a parametric approach. *Int J Syst Sci* 32(4): 533–540.
- Eisenmann, R Sr, Eisenmann, R Jr (1997) Applied condition monitoring. *mach malfunct diagn and correct* 13: 703–741.
- Enbo F, Haibin Y, Ming R (1998) Fuzzy expert system for real-time process condition monitoring and incident prevention. *Expert Syst with Appl* 15: 383–390.
- Eshleman RL (1999) Basic Machinery Vibrations. VIPress Inc, Clarendon Hills, IL.
- Feigl P, Zelen M (1965) Estimation of exponential survival probabilities with concomitant information. *Biom* 21: 826–838.
- Geng Z, Qu L (1994) Vibrational diagnosis of machine parts using the wavelet packet technique. *Insight* 36: 11–15.
- Gideon C (1998) Neural networks implementations to control real-time manufacturing systems. *Comput Integr Manuf Syst* 11: 243–251.
- Gopalan, MN, Kumar D (1995) Analysis of cold-standby systems cutting and clustering the state space. *Int J of Qual Reliab and Saf Eng* 2(3): 327–340.
- Gorzalczany BM (1999) On some idea of a neuro-fuzzy controller. *Inf Sci* 120: 69–87.
- Hans L (1999) Management of industrial maintenance – economic evaluation of maintenance policies. *Int J of Oper & Prod Manag* 19(7): 716–737.

- Hickman *et al.* (1989) Analysis for knowledge-based systems: A practical guide to the KADS methodology. Prentice Hall, Chinchester, UK.
- Husband TM (1978) Maintenance Management and Terotechnology. Gower Publishing, Aldershot, UK.
- Hussain MA (1999) Review of the applications of neural networks in chemical process control-simulation and online implementation. *Artif Intell in Eng* 13: 55–68.
- Isermann R (1993) Fault diagnosis via parameter estimation and knowledge processing. *Autom* 29(4): 815–835.
- Jardine AKS (1987) Maintenance, Replacement and Reliability, Pitman, Boston, MA.
- Jones RB (1995) Risk-based Management: a reliability centered approach. Gulf Publishers, Houston.
- Kazamian BH (2001). Study of Learning Fuzzy Controllers. *Expert Syst*, 18(4): 186–193.
- Kerestecioglu F (1993) Change Detection and Input design in dynamical systems. Research Studies Press, Taunton, UK.
- Kerestecioglu F, Zarrop MB (1994) Input Design for abrupt changes in dynamical systems. *Int J of Control* 59(4): 1063–1084.
- Kinclair LR (1987) Evolution of condition monitoring and the management of maintenance. *Proceedings of International Conference on Condition Monitoring*, 1: 13–21.
- Kitamura M (1980) Detection of sensor failures in nuclear plant using analytic redundancy. *Trans Am Nucl Soc* 34: 581–583.
- Knezevic J (1987) Condition parameter-based approach to calculation of reliability characteristics. *Reliab Eng* 19 (1): 29–39.
- Kobbacy KAH, Fawzy BB, Percy DF (1997) A full history of proportional hazards model for preventive maintenance scheduling. *Qual and Reliab Eng Int* 13: 187–198.
- Kumar U, Granholm S (1990) Reliability centered maintenance – a tool for higher profitability. *Maint* 5(3): 23–26.
- Lau HCW, Wong TT, Ning A (2001) Incorporating machine intelligence in a parameter-based control system: a neural-fuzzy approach. *Artif Intell in Eng* 15(3): 253–264.
- Lebold M, Reichard Karl, Boylan D (2003) Using DCOM in an Open System Architecture Framework for Machinery monitoring and Diagnostics. *IEEE Aerospace Conference*, 3: 1227–1235.
- Leger R P, Garland WJ, Poehlman WFS (1998) Fault detection and diagnosis using statistical control charts and artificial neural networks. *Artif Intell in Eng* 12: 35–47.
- Leung D, Ramanougli J (2000) Dynamic probabilistic model-based expert system for fault diagnosis. *Computers and . Chem Eng* 24: 2473–2492.
- Lin J, Qu L (2000) Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis. *J of Sound and Vib* 234: 135–148.
- Logan D, Joseph M (1994) Using the correlation dimension for the vibration fault diagnosis of rolling element bearings. *Mech Syst and Signal Process* 10(3): 241–250.
- Makis V (1998) Optimal lot sizing and inspection policy for an EMQ model with imperfect inspections. *Nav Res Logist* 45: 165–186.
- Malik MAK (1979) Reliable preventive maintenance scheduling. *AIIE Trans* 11:221–228.
- McFadden PD (1986) Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration. *J of Vib Acoust Stress and Reliab in Des* 108: 165–170.
- McFadden PD (1987) Examination of a technique for early detection of failure in gears by signal processing of the time domain average of the meshing vibration. *Mech Syst and Signal Process* 1: 173–183.
- McFadden PD (1994) Application of the wavelet transform to early detection of gear failure by vibration analysis. *Proceedings of International conference of condition monitoring*, 1, 172–183.

- McFadden PD, Smith JD (1985) A signal processing technique for detecting local defects in a gear from signal average of vibration. *Proceedings of Institute of Mechanical Engineers*, 199(c4): 287–292.
- McFadden PD, Wang WJ (1993) Early detection of gear failure by vibration analysis-I, calculation of the time-frequency distribution. *Mech Syst and Signal Process* 7: 193–203.
- McFadden PD, Wang WJ (1996) Application of wavelets to gearbox vibration signals for fault detection. *J of Sound and Vib* 7: 193–203.
- Mitchell J, Bond T, Bever K, Manning N (1998) MIMOSA Four Years Later. *Sound and Vib* 12–21.
- Mobley RK (1990) *An Introduction to Preventive Maintenance*. Plant Engineering Series, Van Nostrand Reinhold, New York.
- Monderres M (1993) *What every engineer should know about reliability and risk analysis*, M.Dekker, New York.
- Moss MA (1985) *Designing for minimal maintenance expense*. Marcel Drekker Inc, New York.
- Moubray J (1997) *Reliability Centered Maintenance*. Industrial Press, New York.
- Nakajima S (1988) *Total Productive Maintenance*. Productivity Press, Cambridge, Massachusetts.
- Nikoukhah R (1998) Guaranteed active failure detection and isolation for linear dynamical systems. *Autom* 34(11): 348–358
- Nikoukhah R, Campbell SL, Delebecque F (2000) Detection signal design for failure design: A robust approach. *Int J of Adapt Control and Signal Process* 14: 701–724.
- Nooteboom P, Leemeijer GB (1993) Focusing based on the structure of a model in model-based diagnosis. *Intl J Man-Machine Studies* 38: 455–474.
- Ollila A, Malmipuro M (1999) Maintenance has a role in quality. *The TQM Magazine*, 11(1): 17–21.
- Patton RJ (1994) Robust model-based fault diagnosis: the state-of-the-art. *Proceedings of IFAC Symposium on Fault Detection, supervision and Safety for Processes (SAFEPROCESS)*, 1,1:24.
- Patton RJ, Chen J (1994) A review of parity space approaches to fault diagnosis for aerospace systems. *AIAA J of Guid Control & Dyn* 17(2): 278–285.
- Patton RJ, Chen J (1997) Observer-based fault detection and isolation: Robustness and applications. *Control Eng Pract* 5(5): 671–682.
- Paya B, Esat I, Badi MNM (1995) Neural network-based fault detection using different signal processing techniques as pre-processor. *JN Am Soc of Mech Eng PD Publ* 70: 97–101.
- Paya BA, Esat II, Badi MNM (1997) Artificial Neural network-based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mech Syst and Signal Process* 11(5): 751–765.
- Pena E, Hollander M, (1995) Dynamic reliability models with conditional proportional hazards. *Lifetime Data Anal* 1: 377–401.
- Rao BKN (1996) The need for condition monitoring and maintenance management in industries. *Handbook of condition monitoring*, Elsevier Science, Amsterdam, pp 1–36.
- Rao SS (1992) *Reliability-based Design*. McGraw Hill, New York.
- Ray LR, Townsend JR, Ramasubramanian A (2001) Optimal filtering and bayesian detection for friction-based diagnostics in machines. *ISA Trans* 40: 207–221.
- Reiche H (1994) *Maintenance minimization for competitive advantage*. Gordon and Breach Science Publishers, Amsterdam, pp 49–54.
- Reiter R (1987) A theory of diagnosis from first principles. *Artificial Intelligence*, 32: 57–96.

- Sandtorv H (1991) RCM – Closing the loop between design, reliability and operational reliability. *Maint* 6(1): 13–21.
- Saranga H, Knezevic J (2000a) Reliability analysis using multiple relevant condition parameters. *J of Qual in Maint Eng* 6(3): 165–176.
- Saranga H, Knezevic J (2000b) Reliability prediction for condition-based maintained systems. *Reliab Eng Syst Saf* 71, 219–224.
- Sheu C, Krajewski LJ (1994) A decision model for corrective maintenance management. *Int J of Prod Res* 32 (6): 1365–1382.
- Shiroshi J, Li Y, Liang S, Kurfess T, Danyluk S (1997) Bearing Condition Diagnostics via vibration and acoustic emission measurements. *Mech Syst and Signal Process* 11(5): 693–705.
- Simani S, Fantuzii C, Patoon R (2003) *Model-based Fault Diagnosis in Dynamic System Using Identification Techniques*. Springer-Verlag, London, UK.
- Slany W, Vascak J (1996) A consistency checker for a fuzzy diagnosis system applied to warm rolling-mills in steelmaking plants. *Proceedings of the 5th International Conference on Fuzzy Systems*, 1: 206–212.
- Stamatis DH (1995) *Failure Mode and Effect Analysis: FMEA from Theory to Execution*. ASQC Quality Press, Milwaukee, Wisconsin.
- Staszewski WJ (1998) Wavelet-based Compression and feature selection for vibration analysis. *J of Sound and Vib* 211(5): 735–760.
- Staszewski WJ, Tomlinson GR (1994) Application of wavelet transform to fault detection in a spur gear. *Mech Syst and signal Process* 8: 319–356.
- Staszewski WJ, Worden K, Tomlinson GR (1997) Time-Frequency analysis in gearbox fault detection using the wigner-ville distribution and pattern recognition. *Mech Syst and Signal Process* 11(5): 673–692.
- Struss P (1987) *Multiple representation of structure and function*, Expert Systems in Computer Aided Design. Elsevier Science Publishers, Amsterdam.
- Struss P (1988) A framework for model-based diagnosis. Siemens, AG, Technical report, INF ARM-10–88.
- Struss P (1989) Diagnosis as a process. In Hamscher *et al.* (eds). *Readings in model-based diagnosis*, Morgan Kaufman, San Mateo, pp 408–418.
- Struss P, Dressler O (1989) Physical negation: introducing fault models into the General Diagnostic Engine. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, 1318–1323.
- Sun Y (1994) Simulation for maintenance of an FMS: an Integrated system of maintenance and decision making. *IJAMT* 9:35–39.
- Tapiero, CS (1986) Continuous quality production and machine maintenance. *Nav Res Logist Q* 33: 489–499.
- Uosaki K, Tanaka I, Sugiyama H (1984) Optimal Input design for autoregressive model discrimination with constrained output. *IEEE Trans on Automic control* AC-29(4): 348–350.
- Villanueva, H, Lamba H (1997) Operator Guidance System for industrial plant supervision. *Expert Syst with Appl* 12 (4): 441–454.
- Wang CH, Sheung SH (2003) Determining the optimal production-inspection intervals with inspection errors: using a Markov chain. *Comput & Oper Res* 30 (2003): 1–17.
- Wang WJ (2001) Wavelets for detecting mechanical faults with high sensitivity. *Mech Syst and Signal Process* 5(14): 685–696.
- Williams JH, Davies A, Drake PR (1994) *Condition-based maintenance and machine diagnostics*. Chapman & Hall, London, pp 1–18.
- Wilsky AS (1976) A survey of design methods for failure detection in dynamic systems. *Autom* 12(6): 601–611.

- Wireman T (1998) Developing performance indicators for managing maintenance. Industrial Press, New York.
- Won JK, Modarres M (1998) Improved Bayesian method for diagnosing equipment partial failures in process plants. *Comput Chem Eng* 22(10): 1483–1502.
- Wu X, Chen J, Wang W, Zhou Y (2001) Multi-Index Fusion-based Fault diagnosis theories and methods. *Mech Syst and Signal Process* 15(5): 995–1006.
- Yam RCM, Tse PW, Li L, Tu P (2001) Intelligent predictive decision support system for condition-based maintenance. *Int J of Adv Manuf Technol* 17: 383–391.
- Ying H (1998) The takagi-sugeno fuzzy controllers using the simplified linear control rules are nonlinear variable gain controllers. *Autom*, 34(2): 157–167.
- Zhang H, Zhao G (1999) CMEOC – an expert system in the coal mining industry. *Expert Syst with Appl* 16 (1): 73–77.
- Zhang XJ (1989) Auxiliary signal design in fault detection and diagnosis. Springer-Verlag, Hiedelberg, 1989.

Applied Maintenance Models

K. Ito and T. Nakagawa

15.1 Introduction

In the advanced nations, comfortable lives of citizens depend on a wide variety of social infrastructures such as electricity, gas, waterworks, sewerage, traffic, information networks, and so on. For the steady operation of these infrastructures without any serious troubles such as emergency stop of operation, the steady maintenance is indispensable and the maintenance budget becomes extremely expensive in the most advanced nations because of the high personnel costs. In the twenty first century, such conflicts between the needs of utmost variety of infrastructures and the demands of least maintenance budgets becomes a serious social and industrial issue in these nations. Cost-effective maintenance has become an important key technology to resolve the inherent conflict.

Maintenance is classified into preventive maintenance (PM) and corrective maintenance (CM); PM is a maintenance policy in which we undergoe maintenance on a specific schedule before failure, and CM is a maintenance policy which undergoes maintenance after failure (Nakagawa, 2005, 2007; Barlow and Proschan, 1965). Many researchers have studied optimal PM policies because the CM cost at failure is much higher than the PM one and the well-thought-out cost-effective PM can reduce the system maintenance cost dramatically. Detailed investigation of the target system characteristics is required for the consideration of cost-effective PM policy and an individual system which has its peculiar characteristics has its own optimal PM policy.

In this chapter, we consider optimal maintenance models for four different systems: missiles, phased array radar, full authority digital electronic control (FADEC) and co-generation systems based on our original research. Missiles are one of the most representative military systems (Japan Ministry of Defense, 2007). During the Cold War era, the national defense budgets were given priority in most nations. Now in the post-Cold War era, there is no such exceptional priority in the national budget among advanced democratic nations and the maintenance cost of

military systems is designed to be minimum throughout its lifecycle from the primary design phase (US Congress, 1992).

The missile spends almost all of its lifetime in a storage condition and such operational feature of the missile is unique compared with other military and ordinary industrial systems (Bauer *et al.* 1973). During storage condition, missiles degrade gradually and failed missiles cannot be detected except for the function test. That, the function test cannot detect all failed missiles because the environmental condition of flight after launch is extremely severe compared with that of the function test on the ground. These tests are implemented periodically and the optimal test interval must be established which minimizes the maintenance cost and satisfies the required system reliability.

The phased array radar is the latest radar system and its antenna consists of a huge number of uniform tiny antennas (Brookner, 1985). The radar is designed to tolerate a certain amount of failed tiny antennas because the increase of failed tiny antennas degrades the radar performance. Failed tiny antennas cannot be detected during operation and are detected by a function test during intermission of system. The frequent test reduces the system availability and infrequent testing also reduces system availability because the radar cannot sustain the required performance. Therefore, the optimal test interval which maximizes the system availability must be determined.

The FADEC is widely utilized as the fuel controller of gas turbine engines because it can realize a complicated and delicate control compared with a traditional hydro mechanical controller (HMC) (Robinson, 1987). Because the gas turbine engine is a sensitive internal-combustion engine, the rough control may cause serious states such as overspeed and overtemperature, and these result in catastrophic disasters such as the burst of turbine blades and the meltdown of the combustor. As the FADEC of an industrial gas turbine engine system, PLCs (programmable logic controllers) are utilized because they are tiny, have high capacity and are reasonably priced. Gas turbine makers which utilize PLCs for FADECs, have to guarantee the high reliability of FADECs and establish high reliable FADEC systems adopting the redundant design because PLC makers might not guarantee it. The high-performance self-diagnosis of such redundant FADEC system must be initiated.

Finally, the co-generation system is a power plant which can generate electricity and steam simultaneously, and is an application example of the gas turbine engine system (Witte *et al.* 1988). As a power plant resource, the gas turbine engine has superiority compared with other internal-combustion engines because of its tiny size, the cleanliness of its emitted gas and its low vibration. A gas turbine engine is damaged when it is operated, and it has to undergo overhaul forthwith when the cumulative damage is greater than a prespecified level because the safety insurance of engine maker terminates. From the viewpoint of co-generation system users, the overhaul should therefore be implemented at special periods such as the Christmas vacation and the system should be operated without interruption all year round. So, the system user institutes a managerial level which is lower than overhaul level, and the system should undergo overhaul when its cumulative damage exceeds that managerial level. The optimal managerial level which minimizes the operational cost must be considered.

15.2 Missile Maintenance

A system such as missiles is in storage for a long time from delivery to actual usage and has to hold a high mission reliability when it is used. Figure 15.1 shows an example of a service life cycle of missiles (Bauer *et al.* 1973): after a system is transported to each firing operation unit *via* depot, it is installed on a launcher and is stored in a warehouse for a great part of its lifetime, until its operation. A missile is often called a dormant system.

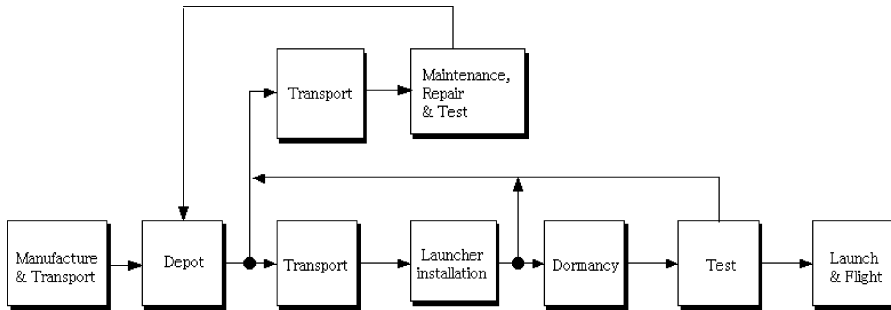


Figure 15.1. A service life cycle of missiles (Bauer *et al.* 1973)

However, the reliability of a storage system goes down with time because some kinds of electronic and electric parts of a system degrade with time (Cottrell *et al.* 1967, 1974; Malik and Mitchell, 1978; Trapp *et al.* 1981; Menke, 1983). For example, Menke confirmed by an accelerated test that integrated circuits of a system might deteriorate in storage condition and it might be impossible to operate when it is necessary (Menke, 1983). Therefore, we should inspect and maintain a storage system at periodic times to hold a high reliability, because it is impossible to inspect whether a storage system can operate normally or not.

Barlow and Proschan summarized optimal inspection policies which minimize the expected cost until detection of failure (Barlow and Proschan, 1965). Zacks and Fenske (1973) and Luss and Kander (1974) extended to a much more complicated system. Shima and Nakagawa (1984) discussed the inspection of a machine with protective devices. Nakagawa (1980) and Thomas *et al.* (1987) considered the inspection policy for a standby unit as an example of a standby electric generator. Martinez (1984) discussed the periodic testing of an electronic equipment in storage for a long period, and showed how to compute its reliability after 10 years of storage.

In the above previous studies, it has been assumed that the function test can clarify all of the kinds of system failures. However, a missile is exposed to a very severe flight environment after launch and some kinds of failures are revealed only in such severe conditions. That is, some failures of a missile cannot be detected by the function test on the ground. To solve this problem, we assume that a system is divided into two independent units: unit 1 becomes new after every inspection because all failures of unit 1 are detected by the function test and are removed completely by maintenance. Meanwhile, unit 2 degrades steadily with time from

delivery to overhaul because all failures of unit 2 cannot be detected by any tests. The reliability of a system deteriorates gradually with time as the reliability of unit 2 deteriorates steadily. A schematic diagram of a missile is given in Figure 15.2.

This section considers a system in storage which is required to have a higher reliability than a prespecified level q ($0 < q \leq 1$) (Ito and Nakagawa, 1992, 1994). To hold reliability, a system is tested and is maintained at periodic times NT ($N = 1, 2, \dots$), and is overhauled if the reliability becomes equal to or lower than q . An inspection number N^* and the time $N^*T + t_0$ until overhaul, are derived when a system reliability is just equal to q . Using them, the expected cost $C(T)$ until overhaul is obtained, and an optimal inspection time T^* which minimizes it is computed. Finally, numerical examples are given when failure times of units have exponential and Weibull distributions.

Further, we consider an extended model where a system consists of unit 1 and units 21 and 22 is partially replaced at N -th inspection (Ito and Nakagawa, 1995). The optimal replacement number M^* which minimizes the expected cost is computed numerically.

15.2.1 Expected Cost

A system consists of units 1 and 2, where unit i has a hazard rate function $H_i(t)$ ($i = 1, 2$). When a system is inspected at periodic times NT ($N = 1, 2, \dots$), unit 1 is maintained and is like new after every inspection, and unit 2 is not done, i.e., its hazard rate remains unchanged by any inspections.

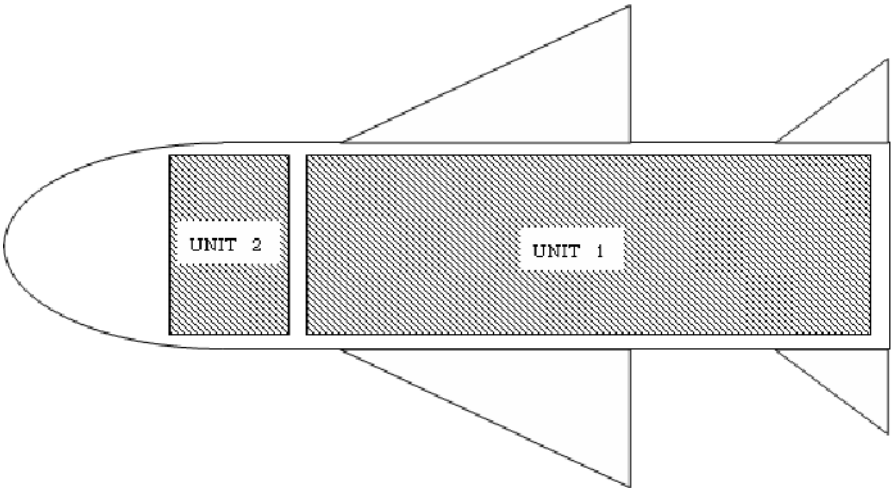


Figure 15.2. Schematic diagram of a missile

From the above assumptions, the reliability function $R(t)$ of a system with no inspection is

$$R(t) = e^{-H_1(t)-H_2(t)} \tag{15.1}$$

If a system is inspected and maintained at time t , the reliability just after the inspection is

$$R(t_{+0}) = e^{-H_2(t)} \quad (15.2)$$

Thus, the reliabilities just before and after the N -th inspection are, respectively,

$$R(NT_{-0}) = e^{-H_1(T)-H_2(NT)} \quad (15.3)$$

and

$$R(NT_{+0}) = e^{-H_2(NT)} \quad (15.4)$$

Next, suppose that the overhaul is performed if the system reliability is equal to or lower than q . Then, if

$$e^{-H_1(T)-H_2(NT)} > q \geq e^{-H_1(T)-H_2[(N+1)T]} \quad (15.5)$$

the time to overhaul is $NT + t_0$, where t_0 ($0 < t_0 \leq T$) satisfies

$$e^{-H_1(t_0)-H_2(NT+t_0)} = q \quad (15.6)$$

This shows that the reliability is greater than q just before the N -th inspection and is equal to q at time $NT + t_0$.

The expected cost per unit time is, from Ross (1970),

$$\frac{\text{Expected cost per cycle}}{\text{Expected time per cycle}}.$$

Defining the time interval $[0, NT + t_0]$ as one cycle, the expected cost until overhaul is given by

$$C(T) = \frac{Nc_1 + c_2}{NT + t_0}, \quad (15.7)$$

where cost c_1 is an inspection cost and c_2 is an overhaul cost.

15.2.2 Optimal Inspection Policies

We consider two particular cases where hazard rate functions $H_i(t)$ are exponential and Weibull ones. An inspection number N^* which satisfies Equation 15.5, and t_0 which satisfies Equation 15.6, are computed. Using these quantities, we compute the expected cost $C(T)$ until overhaul and seek an optimal inspection time T^* which minimizes it.

15.2.2.1 Exponential Case

Suppose that the system obeys an exponential distribution, i.e., $H_i(t) = \lambda_i t$. Then, Equation 15.5 is rewritten as

$$\frac{1}{Na+1} \ln \frac{1}{q} \leq \lambda T < \frac{1}{(N-1)a+1} \ln \frac{1}{q}, \quad (15.8)$$

where

$$\lambda \equiv \lambda_1 + \lambda_2, \quad a \equiv \frac{H_2(T)}{H_1(T) + H_2(T)} = \frac{\lambda_2}{\lambda} \quad (15.9)$$

and a represents an efficiency of inspection, and is adopted widely in practical reliability calculations of storage system (Bauer *et al.* 1973).

When an inspection time T is given, an inspection number N^* which satisfies Equation 15.8 is determined. Particularly, if $\ln 1/q \leq \lambda T$ then $N^* = 0$, and N^* diverges as λT tends to 0. In this case, Equation 15.6 is rewritten as

$$N^* \lambda_2 T + \lambda t_0 = \ln \frac{1}{q}. \quad (15.10)$$

From Equation 15.10, we can compute t_0 easily.

Thus, the total time to overhaul is

$$N^* T + t_0 = N^* (1-a)T + \frac{1}{\lambda} \ln \frac{1}{q}, \quad (15.11)$$

and the expected cost is

$$C(T) = \frac{N^* c_1 + c_2}{N^* (1-a)T + \frac{1}{\lambda} \ln \frac{1}{q}}. \quad (15.12)$$

When an inspection time T is given, we compute N^* from Equation 15.8 and $N^* T + t_0$ from Equation 15.11. Substituting these values into Equation 15.12, we have $C(T)$. Changing T from 0 to $\ln(1/q)/[\lambda(1-a)]$, we can compute an optimal T^* which minimizes $C(T)$. In particular case of $\lambda T \geq \ln(1/q)/(1-a)$, $N^* = 0$ and the expected cost becomes constant, *i.e.*,

$$C(T) = \frac{c_2}{t_0} = -\frac{\lambda c_2}{\ln q}. \quad (15.13)$$

15.2.2.2 Weibull Case

Suppose that the system obeys a Weibull distribution, *i.e.*, $H_i(t) = (\lambda_i t)^m$ ($i=1,2$). Equations 15.5 and 15.6 are rewritten as, respectively,

$$\left\{ \frac{1}{a[(N+1)^m - 1] + 1} \ln \frac{1}{q} \right\}^{\frac{1}{m}} \leq \lambda T < \left[\frac{1}{a(N^m - 1) + 1} \ln \frac{1}{q} \right]^{\frac{1}{m}} \quad (15.14)$$

$$(1-a)t_0^m + a(NT + t_0)^m = \frac{1}{\lambda^m} \ln \frac{1}{q}, \quad (15.15)$$

where

$$\begin{aligned}\lambda^m &\equiv \lambda_1^m + \lambda_2^m, \\ a &\equiv \frac{H_2(T)}{H_1(T) + H_2(T)} = \frac{\lambda_2^m}{\lambda_1^m + \lambda_2^m}.\end{aligned}\quad (15.16)$$

When an inspection time T is given, N^* and t_0 are computed from Equations 15.14 and 15.15. Substituting these values into Equation 15.7, we have $C(T)$, and changing T from 0 to $[\ln(1/q)/(1-a)]^{1/m}/\lambda$, we can compute an optimal T^* which minimizes $C(T)$.

Next, suppose that unit 1 obeys a Weibull distribution with order 2 and unit 2 obeys an exponential distribution, *i.e.*, $H_1(t) = (\lambda_1 t)^2$ and $H_2(t) = \lambda t$. Then, from Equations 15.5 and 15.6, we have, respectively,

$$\begin{aligned}&\frac{1}{2(1-a)^2} \left\{ -(N+1)a + \sqrt{(N+1)^2 a^2 - 4(1-a)^2 \ln q} \right\} \\ &\leq \lambda T < \frac{1}{2(1-a)^2} \left\{ -Na + \sqrt{N^2 a^2 - 4(1-a)^2 \ln q} \right\},\end{aligned}\quad (15.17)$$

$$[(1-a)\lambda t_0]^2 + a\lambda(NT + t_0) + \ln q = 0, \quad (15.18)$$

where a and λ are given in Equation 15.9.

When an inspection time T is given, an inspection number N^* which satisfies Equation 15.17 is computed. Then, the total time to overhaul is

$$\begin{aligned}N^*T + t_0 &= N^*T \\ &+ \frac{1}{2(1-a)^2 \lambda} \left\{ -a + \sqrt{a^2 - 4(1-a)^2 (N^* a \lambda T + \ln q)} \right\}.\end{aligned}\quad (15.19)$$

The expected cost until overhaul is, from Equation 15.7

$$C(T) = \frac{N^* c_1 + c_2}{N^*T + \frac{1}{2(1-a)^2 \lambda} \left\{ -a + \sqrt{a^2 - 4(1-a)^2 (N^* a \lambda T + \ln q)} \right\}} \quad (15.20)$$

In particular, if

$$\lambda T \geq \frac{-a + \sqrt{a^2 - 4(1-a)^2 \ln q}}{2(1-a)^2}, \quad (15.21)$$

then $N^* = 0$ and the expected cost is

$$C(T) = \frac{2(1-a)^2 \lambda c_2}{-a + \sqrt{a^2 - 4(1-a)^2 \ln q}}.$$

(15.22)

Table 15.1. Inspection number N^* and total time to overhaul $\lambda(N^*T^* + t_0)$ for λT when $a = 0.1$ and $q = 0.8$

λT	N^*	$\lambda(N^*T + t_0)$
$[0.223, \infty)$	0	$[0.223, \infty)$
$[0.203, 0.223)$	1	$[0.406, 0.424)$
$[0.186, 0.203)$	2	$[0.558, 0.588)$
$[0.172, 0.186)$	3	$[0.687, 0.725)$
$[0.159, 0.172)$	4	$[0.797, 0.841)$
$[0.149, 0.159)$	5	$[0.893, 0.940)$
$[0.139, 0.149)$	6	$[0.976, 1.026)$
$[0.131, 0.139)$	7	$[1.050, 1.102)$
$[0.124, 0.131)$	8	$[1.116, 1.168)$
$[0.117, 0.124)$	9	$[1.174, 1.227)$
$[0.112, 0.117)$	10	$[1.227, 1.280)$

Therefore, when an inspection time T is given, we compute N^* from Equation 15.17 or Equation 15.21, and $N^*T + t_0$ from Equation 15.19.

Substituting them into Equation 15.20 or Equation 15.22, we compute $C(T)$, and changing T from 0 to $[-a + \sqrt{a^2 - 4(1-a)^2 \ln q}]/[2(1-a)^2 \lambda]$, we can determine T^* which minimizes $C(T)$.

15.2.3 Numerical Illustrations

We specify an algorithm to compute an optimal inspection time T^* :

1. Choose T arbitrarily and seek N^* which satisfies inequalities at Equations 15.8, 15.14 or 15.17.
2. Solve Equations 15.10, 15.15 or 15.18 by Newton-Raphson method, and compute t_0 and $C(T)$ numerically; and
3. Change T and seek an optimal T^* which minimizes $C(T)$ by repeating steps 1 and 2.

Suppose that the failure time of unit i has an exponential distribution $[1-\exp(-\lambda_i t)]$. Table 15.1 gives the optimal inspection number N^* and the total time $\lambda(N^*T + t_0)$ to overhaul for λT when $a = 0.1$ and $q = 0.8$. For example, when λT increases from 0.203 to 0.223, $N^* = 1$ and $\lambda(N^*T + t_0)$ increases from 0.406 to 0.424. In accordance with decreases of λT , both N^* and $\lambda(N^*T + t_0)$ increase as shown in Equations 15.8 and 15.11.

Table 15.2. Optimal inspection time λT^* , total time to overhaul $\lambda(N^*T^* + t_0)$ and minimum expected cost $C(T^*)/\lambda$ when $c_I = I$, $a = 0.1$ and $q = 0.8$

c_2 / c_1	a	q	N^*	λT^*	$\lambda(N^*T^* + t_0)$	$C(T^*) / \lambda$
10	0.1	0.8	8	0.131	1.168	15.41
50	0.1	0.8	19	0.080	1.586	43.51
10	0.5	0.8	2	0.149	0.372	32.27
10	0.1	0.9	7	0.062	0.552	32.63

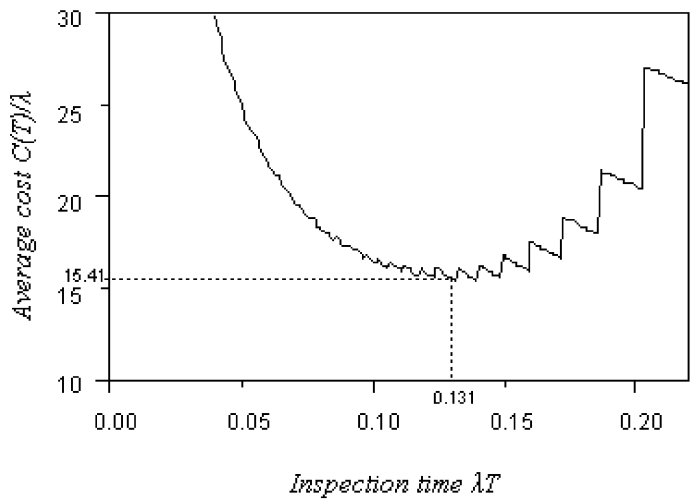


Figure 15.3. Relationship between λT and $C(T)/\lambda$ in exponential case

Table 15.2 gives the optimal inspection number N^* and the optimal time λT^* which minimizes the expected cost $C(T)$ for c_2 / c_1 , a and q , and the resulting total time $\lambda(N^*T + t_0)$ and expected cost $C(T^*)/\lambda$ for $c_1 = 1$. These show that λT^* increases and $\lambda(N^*T + t_0)$ decreases when c_1 / c_2 and a increase, and both λT^* and $\lambda(N^*T + t_0)$ decrease when q increases. Further, Figure 15.3 shows the

relationship between λT and $C(T)/\lambda$ and that the optimal time λT^* which minimizes $C(T)/\lambda$ is 0.131 and the expected cost is 15.41.

Next, suppose that the failure time of unit i has a Weibull distribution $\{1 - \exp[-(\lambda_i t)^m]\}$. When $c_1 = 1$, $c_2 = 10$, $a = 0.1$, $q = 0.8$ and $m = 1.5$, Figure 15.4 shows the relationship λT and $C(T)/\lambda$, and that the optimal time λT^* is 0.230 and the resulting cost $C(T^*)/\lambda$ is 11.19. In this case, the optimal number N^* is 5 and the total time $\lambda(N^*T + t_0)$ is 1.34.

Finally, suppose that the failure time of unit 1 has a Weibull distribution and that of unit 2 has an exponential distribution. Table 15.3 gives the optimal time λT^* and the minimum cost $C(T)/\lambda$ for c_2/c_1 when $c_1 = 1$, $a = 0.1$ and $q = 0.8$. This shows the same tendency as Table 15.2.

If λ is given, we can easily compute the optimal time T^* . For example, when $\lambda = 10^{-5}/\text{h}$ in Figure 15.3, T^* is 1.31×10^4 h, and when $\lambda^{1.5} = 10^{-5}$, i.e., $\lambda = 4.64 \times 10^{-4}/\text{h}$ in Figure 15.4, T^* is 4.96×10^2 . It is expected that T^* decreases when m increases.

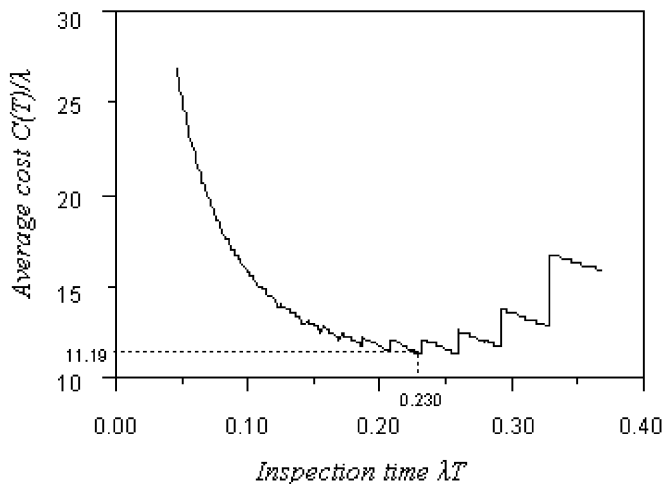


Figure 15.4. Relationship between λT and $C(T)/\lambda$ in Weibull case

Table 15.3. Optimal inspection time λT^* and minimum expected cost $C(T^*)/\lambda$ when $c_1 = 1$, $a = 0.1$ and $q = 0.8$.

c_2/c_1	λT^*	$C(T^*)/\lambda$
10	0.300	8.59
20	0.248	14.01
30	0.227	19.12
50	0.193	28.98

15.3 Phased Array Radar Maintenance

A phased array radar (PAR) is a radar which steers the electromagnetic wave direction electrically. Compared with conventional radars which steer their electromagnetic wave direction by moving their antennas mechanically, a PAR has no mechanical portion to steer its wave direction, and hence it can steer very quickly. Most anti-aircraft missile systems and early warning systems have presently adopted PARs because they can acquire and track multiple targets simultaneously.

A PAR antenna consists of a large number of small and homogeneous element antennas which are arranged flatly and regularly, and steers its electromagnetic wave direction by shifting signal phases of waves which are radiated from these individual elements (Brookner, 1991 and Skolnik, 1980, 1990).

An increase in the number of failed elements degrades the radar performance, and at last, this may cause an undesirable situation such as the omission of targets (Brookner, 1991). The detection, diagnosis, localization and replacement of failed elements of a PAR antenna are indispensable to holding a certain required level of radar performance. A digital computer system controls a whole PAR system, and it detects, diagnoses and localizes failed elements. However, such maintenance actions interrupt the radar operation and decrease its availability. Maintenance interruptions should be minimized. For the above reasons, it would be important to decide an optimal maintenance policy for a PAR antenna, by comparing the downtime loss caused by its maintenance with the degradational loss caused by its performance downgrade.

Recently, a new method of failure detection for PAR antenna elements has been proposed by measuring the electromagnetic wave pattern (Bucci *et al.* 2000). This method could detect some failed elements even when a radar system is operating, *i.e.*, it could be applied to the detection of confined failure modes such as power on-off failures. However, it would be generally necessary to stop the PAR operation for the detection of all failed elements.

Keithley (1966) showed by Monte Carlo simulation that the maintenance time of PAR with 1024 elements had a strong influence on its availability. Hevesh (1967) discussed the following three types of maintenance of PAR in which all failed elements could be detected immediately, and calculated the average times to failures of its equipment, and its availability in immediate maintenance:

- *Immediate maintenance*: failed elements are detected, localized and replaced immediately;
- *Cyclic maintenance*: failed elements are detected, localized and replaced periodically; and
- *Delayed maintenance*: failed elements are detected and localized periodically, and replaced when their number has exceeded a predesignated level.

Further, Hesse (1975) analyzed the field maintenance data of U.S. Army prototype PAR, and clarified that the repair times have a lognormal distribution. In the actual maintenance, the immediate maintenance is rarely adopted because

frequent maintenance degrades a radar system availability. Either cyclic or delayed maintenance is commonly adopted.

We have studied the comparison of cyclic and delayed maintenance of PAR considering the financial optimum (Ito *et al.* 1999; Ito and Nakagawa, 2004). In the study, we derived the expected costs per unit time and discussed the optimal policies which minimize them analytically in these two types of maintenance, and concluded that the delayed maintenance is better than cyclic maintenance in suitable conditions by comparing these two costs numerically. Although the financial optimum takes priority for non-military systems and military systems in the non-combat condition, the operational availability should take more priority than economy for military systems in the combat condition. Therefore, maintenance policies which maximize availability should be considered.

In this section, we perform the periodic detection of failed elements of a PAR where it consists of N_0 elements and failures are detected at scheduled time interval (Nakagawa and Ito, 2007): if the number of failed elements has exceeded a specified number N ($0 < N \leq N_0$), a PAR cannot hold a required level of radar performance, and it causes the operational loss such as the target oversight to a PAR. We assume that failed elements occur at a Poisson process, and consider cyclic, delayed and two modified maintenance schemes. Applying the method of Nakagawa (1986) to such maintenances, availability is obtained, and optimal policies which maximize them are analytically discussed in cyclic and delayed maintenance scenarios. In a numerical example, we decide which type of maintenance is better by comparing availability.

15.3.1 Cyclic Maintenance

We consider the following cyclic maintenance of a PAR (Nakagawa and Ito, 2007):

1. A PAR consists of N_0 elements which are independent and homogeneous on all plains of PAR, and which have an identical constant hazard rate λ_0 . The number of failed elements at time t has a binomial distribution with mean $N_0[1 - \exp(-\lambda_0 t)]$. Since N_0 is large and λ_0 is very small, it might be assumed that failures of elements occur approximately at a Poisson process with mean $\lambda \equiv N_0 \lambda_0$. That is, the probability that j failures occur during $(0, t]$ is

$$P_j(t) \equiv \frac{(\lambda t)^j e^{-\lambda t}}{j!} \quad (j = 0, 1, 2, \dots).$$

2. When the number of failed elements has exceeded a specified number N , a PAR cannot hold a required level of radar performance such as maximum detection range and resolution.
3. Failed elements cannot be detected during operation and can be ascertained only according to the diagnosis software executed by a PAR system computer. Failed elements are usually detected at periodic diagnosis. The

diagnosis is performed at time interval T and a single diagnosis spends time T_0 .

4. All failed elements are replaced by new ones at the M -th diagnosis or at the time when the number of failed elements has exceeded N , whichever occurs first. The replacement spends time T_1 .

When the replacement time of failed element antennas is assumed to be the regeneration point, the availability of system is denoted as

$$A = \frac{\text{Effective time between regeneration points}}{\text{Total time between regeneration points}}. \quad (15.23)$$

When the number of failed elements is below N at the M -th diagnosis, the expected effective time until replacement is

$$MT \sum_{j=0}^{N-1} p_j (MT). \quad (15.24)$$

When the number of failed elements exceeds N at the i ($i = 1, 2, \dots, M$)-th diagnosis, the expected effective time until replacement is

$$\sum_{i=1}^M \sum_{j=0}^{N-1} p_j [(i-1)T] \sum_{k=N-j}^{\infty} \int_{(i-1)T}^{iT} t dp_k [t - (i-1)T]. \quad (15.25)$$

Thus, from Equations 15.24 and 15.25, the total expected effective time until replacement is

$$T \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j (iT) - \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j (iT) \sum_{k=N-j+1}^{\infty} \frac{k-N+j}{\lambda} p_k (T). \quad (15.26)$$

Next, when the number of failed elements is below N at the M -th diagnosis, the expected time between two adjacent regeneration points is

$$\sum_{j=0}^{N-1} p_j (MT) [M(T + T_0) + T_1]. \quad (15.27)$$

When the number of failed elements exceeds N at the i ($i = 1, 2, \dots, M$)-th diagnosis, the expected time between two adjacent regeneration points is

$$\sum_{i=1}^M \sum_{j=0}^{N-1} p_j [(i-1)T] \sum_{k=N-j}^{\infty} [i(T + T_0) + T_1] p_k (T). \quad (15.28)$$

Thus, from Equations 15.27 and 15.28, the total expected time between two adjacent regeneration points is

$$T_1 + (T + T_0) \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j (iT). \quad (15.28)$$

Therefore, by dividing Equation 15.26 by Equation 15.29, the availability of cyclic maintenance $A_1(M)$ is

$$\begin{aligned}
A_1(M) &= \frac{T \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j(iT) - \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j(iT) \sum_{k=N-j+1}^{\infty} (k-N+j) p_k(T) / \lambda}{T_1 + (T+T_0) \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j(iT)} \\
&= \frac{T}{T+T_0} \left[1 - \frac{\begin{aligned} &T_1/(T+T_0) \\ &+ \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j(iT) \\ &\times \sum_{k=N-j+1}^{\infty} (k-N+j) p_k(T) / (\lambda T) \end{aligned}}{\frac{T_1}{T+T_0} + \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j(iT)} \right] \\
&= \frac{T}{T+T_0} [1 - \bar{A}_1(M)] \quad (M = 1, 2, \dots, \infty) \quad (15.30)
\end{aligned}$$

Because maximizing availability $A_1(M)$ is equal to minimizing unavailability $\bar{A}_1(M)$ from Equation 15.30, we consider M^* which minimizes unavailability $\bar{A}_1(M)$. Forming the inequality $\bar{A}_1(M+1) - \bar{A}_1(M) \geq 0$,

$$\begin{aligned}
L_1(M) &\left[\frac{T}{T+T_0} + \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j(iT) \right] \\
&- \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p_j(iT) \sum_{k=N-j+1}^{\infty} \frac{k-N+j}{\lambda} p_k(T) \geq \frac{T_1 T}{T+T_0} \quad (15.31)
\end{aligned}$$

where

$$L_1(M) \equiv \frac{\sum_{j=0}^{N-1} p_j(MT) \sum_{k=N-j+1}^{\infty} (k-N+j) p_k(T) / \lambda}{\sum_{j=0}^{N-1} p_j(MT)}. \quad (15.32)$$

Let $Q_1(M)$ denote the left-hand side of Equation 15.31. Then,

$$\begin{aligned}
&Q_1(M+1) - Q_1(M) \\
&= [L_1(M+1) - L_1(M)] \left[\frac{T_1}{T+T_0} + \sum_{i=0}^M \sum_{j=0}^{N-1} p_j(iT) \right], \quad (15.33)
\end{aligned}$$

Thus, if $L_1(M)$ is strictly increasing in M , then $Q_1(M)$ is strictly increasing in M . Therefore, we have the following optimal policy.

Theorem 15.3.1

1. If $L_1(M)$ is strictly increasing in M and $Q_1(\infty) > T_1 T / (T + T_0)$ then there exists a finite and unique M^* which satisfies Equation 15.31.
2. If $L_1(M)$ is strictly increasing in M and $Q_1(\infty) \leq T_1 T / (T + T_0)$ then $M^* = \infty$;
3. If $L_1(M)$ is decreasing in M then $M^* = 1$ or $M^* = \infty$.

15.3.2 Delayed Maintenance

We consider the delayed maintenance of a PAR (Nakagawa and Ito, 2007). All failed elements are replaced by new ones only when failed elements have exceeded a managerial number $N_c (< N)$ at diagnosis. The replacement spends time T_1 . The other assumptions are the same as those in Section 15.3.1.

When the number of failed elements is between N_c and N , the expected effective time until replacement is

$$\sum_{i=1}^{\infty} \sum_{j=0}^{N_c-1} p_j [(i-1)T] \sum_{k=N_c-j}^{N-j-1} i T p_k(T). \quad (15.34)$$

When the number of failed elements exceed N , the expected effective time until replacement is

$$\sum_{i=1}^{\infty} \sum_{j=0}^{N_c-1} p_j [(i-1)T] \sum_{k=N-j}^{\infty} \int_{(i-1)T}^{iT} t dp_k[t - (i-1)T]. \quad (15.35)$$

Thus, the total expected effective time until replacement is, from Equations 15.34 and 15.35,

$$T \sum_{i=0}^{\infty} \sum_{j=0}^{N_c-1} p_j (iT) - \sum_{i=0}^{\infty} \sum_{j=0}^{N_c-1} p_j (iT) \sum_{k=N-j+1}^{\infty} \frac{k - N + j}{\lambda} p_k(T). \quad (15.36)$$

Similarly, when the number of failed elements is between N_c and N , the expected time between two adjacent regeneration points is

$$\sum_{i=1}^{\infty} \sum_{j=0}^{N_c-1} p_j [(i-1)T] \sum_{k=N_c-j}^{N-j-1} [i(T + T_0) + T_1] p_k(T). \quad (15.37)$$

When the number of failed elements exceeds N , the expected time between two adjacent regeneration points is

$$\sum_{i=1}^{\infty} \sum_{j=0}^{N_c-1} p_j [(i-1)T] \sum_{k=N-j}^{\infty} [i(T + T_0) + T_1] p_k(T). \quad (15.38)$$

Thus, the total expected time between two adjacent regeneration points is, from Equations 15.37 and 15.38,

$$(T + T_0) \sum_{i=0}^{\infty} \sum_{j=0}^{N_c-1} p_j(iT) + T_1. \quad (15.39)$$

Therefore, the availability of delayed maintenance $A_2(N_c)$ is, by dividing Equation 15.36 by Equation 15.39,

$$A_2(N_c) = \frac{T}{T + T_0} \left[1 - \frac{\begin{aligned} &T_1/(T + T_0) \\ &+ \sum_{i=0}^{\infty} \sum_{j=0}^{N_c-1} p_j(iT) \\ &\times \sum_{k=N-j}^{\infty} [1 - (N - j)/(k + 1)] p_k(T) \end{aligned}}{T_1/(T + T_0) + \sum_{i=0}^{\infty} \sum_{j=0}^{N_c-1} p_j(iT)} \right] \\ \equiv \frac{T}{T + T_0} [1 - \overline{A_2}(N_c)] \quad (15.40)$$

Forming the inequality $\overline{A_2}(N_c + 1) - \overline{A_2}(N_c) \geq 0$, we have

$$\frac{E_{N_c}}{D_{N_c} - E_{N_c}} \sum_{j=0}^{N_c-1} (D_j - E_j) - \sum_{j=0}^{N_c-1} E_j \geq \frac{T_1}{T + T_0}. \quad (15.41)$$

where $D_j \equiv \sum_{i=0}^{\infty} p_j(iT)$ and $E_j \equiv \sum_{i=0}^{\infty} p_j(iT) \sum_{k=N-j}^{\infty} [1 - (N - j)/(k + 1)] p_k(T)$.

Let $Q_2(N_c)$ denote the left-hand side of Equation 15.41 and $L_2(N_c) \equiv E_{N_c} / (D_{N_c} - E_{N_c})$. Then,

$$Q_2(N_c + 1) - Q_2(N_c) = [L_2(N_c + 1) - L_2(N_c)] \sum_{j=0}^{N_c} (D_j - E_j) \quad (15.42)$$

As $D_j - E_j > 0$, the sign of $Q_2(N_c + 1) - Q_2(N_c)$ depends on $L_2(N_c + 1) - L_2(N_c)$. Therefore, we have the following optimal policy:

Theorem 15.3.2

- If $L_2(N_c)$ is strictly increasing in N_c and $Q_2(N) > T_1/(T + T_0)$ then there exists a finite and unique $N_c^* (1 \leq N_c^* \leq N)$ which satisfies Equation 15.41; and

- If $L_2(N_c)$ is strictly increasing in N_c and $Q_2(N) > T_1/(T+T_0)$ then $N_c^* = N$, i.e., the planned maintenance should not be done.

Table 15.4. Optimal number of diagnosis M^* , optimal managerial number of failed elements N_c^* and unavailability $\bar{A}_1(M^*)$ and $\bar{A}_2(N_c^*)$

N	T	T_0	T_1	λ	M^*	$\bar{A}_1(M^*)$	N_c^*	$\bar{A}_2(N_c^*)$
100	24×7	1	8	0.1	5	1.141×10^{-2}	78	0.936×10^{-2}
90	24×7	1	8	0.1	4	1.186×10^{-2}	68	1.058×10^{-2}
70	24×7	1	8	0.1	3	1.574×10^{-2}	49	1.430×10^{-2}
100	24×10	1	8	0.1	3	1.097×10^{-2}	70	0.998×10^{-2}
100	24×14	1	8	0.1	2	1.173×10^{-2}	60	1.113×10^{-2}
100	24×7	0.5	8	0.1	5	1.144×10^{-2}	78	0.938×10^{-2}
100	24×7	0.1	8	0.1	5	1.146×10^{-2}	78	0.941×10^{-2}
100	24×7	1	5	0.1	4	0.735×10^{-2}	77	0.593×10^{-2}
100	24×7	1	2	0.1	4	0.295×10^{-2}	75	0.242×10^{-2}
100	24×7	1	8	0.2	2	2.312×10^{-2}	62	2.143×10^{-2}
100	24×7	1	8	0.3	1	4.520×10^{-2}	48	3.830×10^{-2}

15.3.3 Numerical Illustrations

Table 15.4 gives the optimal number of diagnosis M^* and the optimal managerial number of failed elements N_c^* , and the unavailability $\bar{A}_1(M^*)$ and $\bar{A}_2(N_c^*)$ for $N = 70, 90, 100$, $T = 168, 240, 360$ h (7, 10, 14 days), $T_0 = 0.1, 0.5, 1$, $T_1 = 2, 5, 8$ and $\lambda = 0.1, 0.2, 0.3$ /hours. In all cases in Table 15.1, $L_1(M)$ is strictly increasing in M .

Table 15.4 indicates that M^* and N_c^* decrease when N , $1/T$, T_1 , and $1/\lambda$ decrease, and the change of T_0 hardly affects M^* and N_c^* . In this calculation,

$\bar{A}_1(M^*)$ is always greater than $\bar{A}_2(N_c^*)$. Therefore, we can adjudge in this case that the delayed maintenance provides more availability than the cyclic maintenance.

15.4 Self-diagnosis for FADEC

The original idea of gas turbine engines was represented by Barber in England in 1791, and they were firstly realized in the twentieth century. After that, they advanced greatly during World WarII. Today, gas turbine engines have been widely utilized as the main engines of airplanes, high performance mechanical pumps, emergency generators and cogeneration systems because they can generate high power for their sizes, their start times are very short, and no coolant water is necessary for operation (Robinson, 1987; Kendell, 1981).

Gas turbine engines are mainly constituted with three parts, *i.e.*, compressor, combustor and turbine. The engine control is performed by governing the fuel flow to the engine. When gas turbine engines are operating, dangerous phenomena, such as surge, stool and over-temperature of exhaust gas, should be paid attention to because they may cause serious damage to the engine. To prevent them, the turbine speed, inlet temperature and pressure, and exhaust gas temperature of gas turbine engines are monitored, and an engine controller should determine appropriate fuel flow by checking these data.

The gas turbine engine has to operate in a serious environment and a hydro mechanical controller (HMC) was adopted as the most common engine controller for a long period because of its high reliability, durability and excellent operational response. However, the performance of gas turbine engines has advanced and customers have found a need to decrease the operation cost. HMC could not meet these advanced demands and so the engine controller has been electrified. The first electric engine controller, which was a support unit of HMS, was adopted for the J47-17 turbo jet engine of the F86D fighter in the late 1940s. The evolution of devices, from vacuum tube to transistor and transistor to IC, has changed the role of electric engine controllers from the assistant of HMS to the full authority controller because of the increase in reliability. In the 1960s, the analogue full authority controller could not meet the accuracy demands of engines, and the full authority digital engine controller (FADEC) was developed (Robinson, 1987; Kendell, 1981; Scoles, 1986).

FADEC is an electric engine controller which can perform the complicated signal processes involved with digitized engine data. Aircraft FADECs must generally build up duplicated and triplicated systems because they are expected to provide high mission reliability and are needed to decrease weight, hardware complication and electric consumption (Eccles *et al.* 1980; Davies *et al.* 1983; Cahill and Underwood, 1987). Industrial gas turbine engines have introduced advanced technologies which were established for aircraft engines. FADECs, which were originally developed for aircraft, have now also been adopted for industrial gas turbine engines. Comparing general industrial gas turbine FADECs and aircraft FADECs, the following differences are recognized:

- Aircraft gas turbine FADECs have to perform high-speed data processing because the rapid response for aircraft body movement is necessary and inlet pressure and temperature change greatly depending on flying altitude. On the other hand, industrial gas turbine FADECs are not required such high performance compared to aircraft ones because they operate at steady speed on ground.
- Aircraft gas turbine FADECs have to be reliable and fault tolerable, and therefore, they adopt duplicated and triplicated systems because their malfunction in operation may cause serious damage to aircraft and crews. Industrial gas turbine FADECs also have to be reliable and fault tolerant and still be low cost because they have to be competitive in the market.

Depending on the advance of microelectronics, small, high performance, low cost programmable logic controllers (PLC) have been widely distributed in the market. They were originally developed as the substitute for bulky electric relay logic sequencers of industrial automatic systems. Applying the numerical calculation ability of microprocessors, and analogue-digital and digital-analogue converters, these PLCs can perform numerical control. Appropriating such PLCs, very high performance and low cost FADEC systems can be realized. However, these PLCs are developed as general industrial controllers and PLC makers might not permit them for applying high pressurized and hot fluid controllers. Then gas turbine makers who apply these PLCs to FADECs have to design some protective mechanism and have to assure high reliability.

In this section, we consider self-diagnosis policies for dual, triple and N redundant gas turbine engine FADECs, and discuss the diagnosis intervals (Ito and Nakagawa, 2003).

15.4.1 Double Module System

Consider the following self diagnosis policy for a hot standby double module FADEC system: Figure 15.5 illustrates an example of the FADEC construction:

1. The FADEC system consists of two independent channels and reliabilities of channel i at time t are $\bar{F}_i(t) (i=1,2)$.
2. The control calculation of each channel is performed at time interval T_0 , and the self-diagnosis and cross-diagnosis are performed synchronously between two channels at every n -th calculation. The coverage of these diagnoses is 100%.
3. When the number n decreases, the diagnosis calculation per unit time increases and degrades the quality of control. It is assumed that the degradation of control is represented as $c_1/(n+T_1)$, where c_1 is constant and T_1 is the percentage of diagnosis time divided by T_0 .
4. When n increases, the time interval from occurrence of failure to its detection is prolonged and it causes the damage of the gas turbine engine because the extraordinary fuel control signal may incur overspeed or overtemperature of engines. The engine damage is represented as

$c_2(nT_0 - t)$, where t is the time that failure occurs and c_2 is the system loss per unit time.

5. Initially, channel 1 is active and channel 2 is in hot standby. When channel 1 fails, it changes to standby and channel 2 changes to active if it does not fail. It is assumed that these elapsed times for changing are negligible. When both channels 1 and 2 have failed, the system makes an emergency stop.

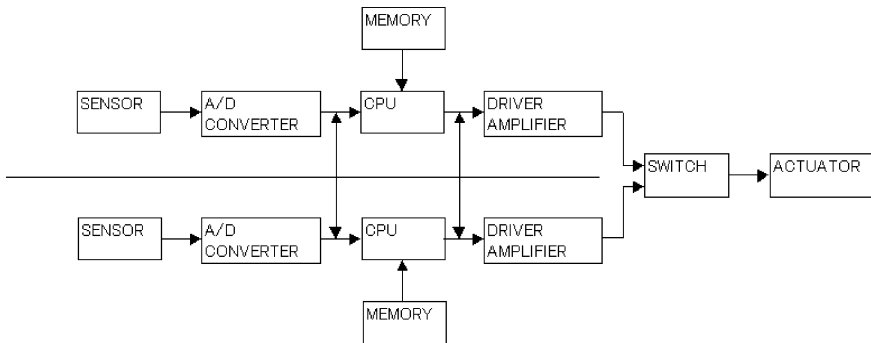


Figure 15.5. Example of double module FADEC construction

When channel i fails at time $t_i (i = 1, 2)$, the following two mean times from failure to its detection are considered:

- When $t_2 \leq t_1 < t_m$ or $t_{m-1} < t_1 < t_2 \leq t_m$, the mean time is

$$\sum_{m=1}^{\infty} F_2(t_m) \int_{t_{m-1}}^t (t_m - t_1) dF_1(t_1), \quad (15.43)$$

where $t_m = mnT_0 (m = 1, 2, 3, \dots)$.

- When $t_1 \leq t_{m-1} < t_2 \leq t_m$, the mean time is

$$\sum_{m=2}^{\infty} \sum_{k=1}^{m-1} \int_{t_{k-1}}^{t_k} (t_k - t_1 + t_m - t_2) dF_1(t_1) \int_{t_{m-1}}^{t_m} dF_2(t_2). \quad (15.44)$$

The total mean time from failure to its detection is the summation of Equations 15.43 and 15.44, and is given by

$$\sum_{m=0}^{\infty} \left\{ \int_{t_m}^{t_{m+1}} [F_1(t) - F_1(t_m)] dt + F_1(t_m) \int_{t_m}^{t_{m+1}} [F_2(t) - F_2(t_m)] dt \right\} \quad (15.45)$$

where $F_1(0) = 0$. Thus, the total expected cost of dual redundant FADEC until the system stops is

$$C_2(n) = \frac{c_1}{n + T_1} + c_2 \sum_{m=0}^{\infty} \left\{ \int_{t_m}^{t_{m+1}} [F_1(t) - F_1(t_m)] dt + F_1(t_m) \int_{t_m}^{t_{m+1}} [F_2(t) - F_2(t_m)] dt \right\} \quad (5.46)$$

Assuming $F_i(t) = 1 - \exp(-\lambda_i t)$ ($i = 1, 2$), Equation (5.46) is rewritten as

$$C_2(n) = \frac{c_1}{n + T_1} + c_2 \left\{ nT_0 \left[\frac{1}{1 - e^{-\lambda_1 nT_0}} + \frac{1}{1 - e^{-\lambda_2 nT_0}} - \frac{1}{1 - e^{-(\lambda_1 + \lambda_2)nT_0}} \right] + \frac{1 - e^{-\lambda_2 nT_0}}{\lambda_2 (1 - e^{-(\lambda_1 + \lambda_2)nT_0})} - \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right\} \quad (15.47)$$

We easily find that

$$C_2(0) = \frac{c_1}{T_1}, \quad C_2(\infty) \equiv \lim_{n \rightarrow \infty} C_2(n) = \infty \quad (15.48)$$

Therefore, there exists a finite n_2^* ($< \infty$) which minimizes $C_2(n)$.

When $\lambda_1 = \lambda_2 \equiv \lambda$, Equation 15.47 is rewritten as

$$C_2(0) = \frac{c_1}{n + T_1} + c_2 \left(\frac{2}{1 - e^{-n\lambda T_0}} - \frac{1}{1 - e^{-2n\lambda T_0}} \right) \left(nT_0 - \frac{1 - e^{-n\lambda T_0}}{\lambda} \right) \quad (15.49)$$

Supposing $x = nT_0$ and C_2 is a continuous function of x , Equation 15.49 is rewritten as

$$C_2(x) = \frac{c_1 T_0}{x + T_1 T_0} + c_2 \left(\frac{2}{1 - e^{-\lambda x}} - \frac{1}{1 - e^{-2\lambda x}} \right) \left(x - \frac{1 - e^{-\lambda x}}{\lambda} \right) \quad (15.50)$$

Differentiating $C_2(x)$ with respect to x , and putting it to zero, we have

$$\left(\frac{2}{1 - e^{-\lambda x}} - \frac{1}{1 - e^{-2\lambda x}} \right) (1 - e^{-\lambda x}) - 2 \left[\frac{e^{-\lambda x}}{(1 - e^{-\lambda x})^2} - \frac{e^{-2\lambda x}}{(1 - e^{-2\lambda x})^2} \right] (\lambda x - 1 + e^{-\lambda x}) = \frac{c_1 T_0}{c_2 (x + T_1 T_0)^2} \quad (15.51)$$

We compute optimal x^* which minimizes $C_2(x)$, and using these values, we can obtain optimal n^* which minimizes $C_2(n)$ in Equation 15.49.

15.4.2 Triple Module System

Next, consider the self diagnosis policy for a hot standby triple module FADEC system: we make following assumptions instead of 1. and 5. in Section 15.4.1, respectively:

1. The FADEC system consists of three independent channels and the reliabilities of channel i at time t are $\bar{F}_i(t) (i = 1, 2, 3)$.
2. Initially, channel 1 is active and channels 2 and 3 are in hot standby. When channel 1 fails, channel 1 changes to standby and channel 2 changes to active if it does not fail. Furthermore, when both channels 1 and 2 have failed, they change to standby and channel 3 changes to active if it does not fail. It is assumed that these elapsed times for changing are negligible. When all channels 1, 2 and 3 have failed, the system makes an emergency stop.

When channel i fails at time $t_i (i = 1, 2, 3)$, the following four mean times from failure to its detection are considered:

- When $t_2, t_3 \leq t_1$ or $t_{m-1} < t_1 < t_2 < t_3 \leq t_m$, the mean time is

$$\sum_{m=1}^{\infty} F_2(t_m) F_3(t_m) \int_{t_{m-1}}^{t_m} (t_m - t_1) dF_1(t_1) \quad (15.52)$$

- When $t_{m-1} < t_1 < t_2 \leq t_m < t_3$ or $t_2 \leq t_1 \leq t_m < t_3$, the mean time is

$$\sum_{m=2}^{\infty} \sum_{l=1}^{m-1} F_2(t_l) \int_{t_{l-1}}^{t_l} (t_l - t_1 + t_m - t_3) dF_1(t_1) \int_{t_{m-1}}^{t_m} dF_3(t_3) \quad (15.53)$$

- When $t_{m-1} < t_1 < t_3 \leq t_m < t_2$, $t_1 < t_m < t_2 < t_3 \leq t_{m+1}$, or $t_1 < t_m < t_3 < t_2$, the mean time is

$$\sum_{m=2}^{\infty} \sum_{l=1}^{m-1} F_3(t_m) \int_{t_{l-1}}^{t_l} (t_l - t_1 + t_m - t_2) dF_1(t_1) \int_{t_{m-1}}^{t_m} dF_2(t_2) \quad (15.54)$$

- When $t_1 < t_m < t_2 < t_{m+1} < t_3$, the mean time is

$$\sum_{m=3}^{\infty} \sum_{l=2}^{m-1} \sum_{k=1}^{l-1} \int_{t_{l-1}}^{t_l} (t_k - t_1 + t_l - t_2 + t_m - t_3) dF_1(t_1) \int_{t_{l-1}}^{t_l} dF_2(t_2) \int_{t_{m-1}}^{t_m} F_3(t_m) \quad (15.55)$$

The total mean time is the summation of Equations 15.52–15.55 and is given by

$$\sum_{m=0}^{\infty} \left\{ \int_{t_m}^{t_{m+1}} [F_1(t) - F_1(t_m)] dt + F_1(t_m) \int_{t_m}^{t_{m+1}} [F_2(t) - F_2(t_m)] dt \right. \\ \left. + F_1(t_m) F_2(t_m) \int_{t_m}^{t_{m+1}} [F_3(t) - F_3(t_m)] dt \right\} \quad (15.56)$$

Thus, the total expected cost of triple redundant FADEC until the system stops is

$$C_3(n) = \frac{c_1}{n + T_1} + c_2 \sum_{m=0}^{\infty} \left\{ \int_{t_m}^{t_{m+1}} [F_1(t) - F_1(t_m)] dt \right. \\ \left. + F_1(t_m) \int_{t_m}^{t_{m+1}} [F_2(t) - F_2(t_m)] dt \right. \\ \left. + F_1(t_m) F_2(t_m) \int_{t_m}^{t_{m+1}} [F_3(t) - F_3(t_m)] dt \right\} \quad (15.57)$$

Assuming $F_i(t) = 1 - \exp(-\lambda_i t)$ ($i = 1, 2, 3$) and $\lambda_1 = \lambda_2 = \lambda_3 \equiv \lambda$, Equation 15.57 is rewritten as

$$C_3(n) = \frac{c_1}{n + T_1} + c_2 \left(\frac{3}{1 - e^{-n\lambda T_0}} - \frac{3}{1 - e^{-2n\lambda T_0}} + \frac{1}{1 - e^{-3n\lambda T_0}} \right) \\ \times \left(nT_0 - \frac{1 - e^{-n\lambda T_0}}{\lambda} \right) \quad (15.58)$$

We easily find that

$$C_3(0) = \frac{c_1}{T_1}, \quad C_3(\infty) \equiv \lim_{n \rightarrow \infty} C_3(n) = \infty \quad (15.59)$$

Therefore, there exists a finite $n_3^* (< \infty)$ which minimizes $C_3(n)$.

Supposing $x = nT_0$ and C_3 is a continuous function of x , Equation 15.5) is rewritten as

$$C_3(x) = \frac{c_1 T_0}{x + T_1 T_0} \\ + c_2 \left(\frac{3}{1 - e^{-\lambda x}} - \frac{3}{1 - e^{-2\lambda x}} + \frac{1}{1 - e^{-3\lambda x}} \right) \left(x - \frac{1 - e^{-\lambda x}}{\lambda} \right) \quad (15.60)$$

Differentiating $C_3(x)$ with respect to x and putting it to zero, we have

$$\left(\frac{3}{1 - e^{-\lambda x}} - \frac{3}{1 - e^{-2\lambda x}} + \frac{1}{1 - e^{-3\lambda x}} \right) (1 - e^{-\lambda x}) \\ - 3 \left[\frac{e^{-\lambda x}}{(1 - e^{-\lambda x})^2} - \frac{2e^{-2\lambda x}}{(1 - e^{-2\lambda x})^2} + \frac{e^{-3\lambda x}}{(1 - e^{-3\lambda x})^2} \right] (\lambda x - 1 + e^{-\lambda x}) \\ = \frac{c_1 T_0}{c_2 (x + T_0 T_1)^2} \quad (15.61)$$

We can compute optimal n_3^* which minimizes $C_3(n)$ using x^* which satisfies Equation 15.61.

15.4.3 N Module System

Consider a hot standby N module FADEC system, a FADEC system which consists of N independent channels and channel i has the reliability $\overline{F}_i(t) (i=1,2,\dots,N)$. By the similar method of Sections 15.4.1 and 15.4.2, the total expected cost is

$$\begin{aligned}
 C_N(n) &= \frac{c_1}{n+T_1} + c_2 \sum_{m=0}^{\infty} \left\{ \int_{t_m}^{t_{m+1}} [F_1(t) - F_1(t_m)] dt \right. \\
 &\quad + F_1(t_m) \int_{t_m}^{t_{m+1}} [F_2(t) - F_2(t_m)] dt + \dots \\
 &\quad \left. \dots + F_1(t_m) F_2(t_m) \dots F_{N-1}(t_m) \int_{t_m}^{t_{m+1}} [F_N(t) - F_N(t_m)] dt \right\} \\
 &= \frac{c_1}{n+T_1} \\
 &\quad + c_2 \sum_{m=0}^{\infty} \sum_{k=1}^N F_1(t_m) \dots F_{k-1}(t_m) \int_{t_m}^{t_{m+1}} [F_k(t) - F_k(t_m)] dt,
 \end{aligned} \tag{15.62}$$

where $F_0 \equiv 1$. The total expected costs $C_N(n)$ agree with Equations 15.46 and (15.57) for $N=2,3$, respectively.

15.4.4 Numerical Illustrations

Table 15.5 gives the optimal x^* , n^* and $C(n^*)$ for $c_2/c_1 = 1, 2, 3, 4$; $T_1 = 0.1, 0.2, 0.3, 0.4$ and $\lambda = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$ when $c_1 = 1$, $T_0 = 10^{-2} s = (10 \text{ ms})$.

It is very natural that when c_2/c_1 increases, optimal diagnosis intervals n^* become short, and the total expected costs $C(n^*)$ become high for both double and triple module systems. In this case, when T_1 increases, optimal diagnosis intervals n^* becomes shorter slightly, and the total expected cost $C(n^*)$ become lower barely, for both systems. While, the changing of λ shows no effect on n^* and $C(n^*)$ for double and triple module systems. Comparing double and triple systems, n_2^* is longer and $C_2(n_2^*)$ is lower than n_3^* and $C_3(n_3^*)$, respectively.

Table 15.5. Optimal diagnosis intervals x^* and n^* which minimize the total expected cost $C(n^*)$ of duplicated and triplicated redundant FADECs

c_2 / c_1	T_1	λ	x_2^*	n_2^*	$C_2(n_2^*)$	x_3^*	n_3^*	$C_3(n_3^*)$
1	0.1	10^{-5}	0.1145	11	0.1726	0.1035	10	0.1907
2	0.1	10^{-5}	0.0807	8	0.2435	0.0729	7	0.2692
3	0.1	10^{-5}	0.0657	7	0.2989	0.0593	6	0.3336
4	0.1	10^{-5}	0.0567	6	0.3461	0.0512	5	0.3795
1	0.2	10^{-5}	0.1135	11	0.1718	0.1025	10	0.1897
1	0.3	10^{-5}	0.1125	11	0.1710	0.1015	10	0.1888
1	0.4	10^{-5}	0.1115	11	0.1702	0.1004	10	0.1878
1	0.1	10^{-5}	0.1145	11	0.1726	0.1035	10	0.1907
1	0.1	10^{-5}	0.1145	11	0.1726	0.1035	10	0.1907
1	0.1	10^{-5}	0.1145	11	0.1726	0.1035	10	0.1907

15.5 Co-generation System Maintenance

A co-generation system produces coincidentally both electric power and process heat in a single integrated system, and today is exploited as a widely distributed power plant (Witte *et al.* 1988). Various kinds of generators, such as steam turbine, gas turbine engines, gas engines, and diesel engines, are adopted as the power sources of co-generation systems. A gas turbine engine has some attractive advantages as compared with other power sources, because its size is the smallest, its exhaust gas emission is the cleanest, and both its noise and its vibration level are the lowest in all power sources of the same power output. So gas turbine co-generation systems are now widely utilized in factories, hospitals, and intelligent buildings to reduce costs of fuel and electricity. A schematic diagram of gas turbine engine co-generation system is shown in Figure 15.6.

Maintenance is essential to uphold system availability; however, its cost may oppress customers financially. System suppliers should propose an effective maintenance plan to minimize the financial load on customers. Because the maintenance cost of the gas turbine engine is most of the maintenance cost of a whole system, an efficient maintenance policy should be established.

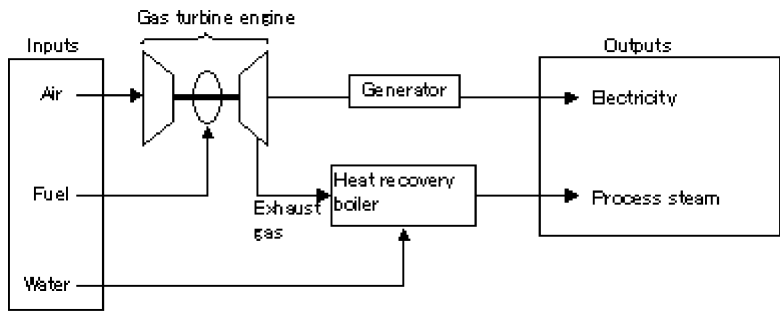


Figure 15.6. Schematic diagram of gas turbine co-generation system

Cumulative damage models have been proposed by many authors (Boland and Proschan, 1983; Cox, 1962; Esary *et al.* 1973; Feldman, 1976, 1977; A-Hameed and Proschan, 1973; A-Hameed and Shimi, 1978; Posner and Zuckerman, 1984; Puri and Singh, 1986; Taylor, 1975; Zuckerman, 1977, 1980). In this section, we discuss the maintenance plan of a gas turbine engine utilizing cumulative damage models (Ito and Nakagawa, 2006). The engine is overhauled when its cumulative damage exceeds a managerial damage level. The expected cost per unit time is obtained and an optimal damage level which minimizes it is derived. Numerical examples are given to illustrate the results.

15.5.1 Model and Assumptions

Customers have to operate their co-generation system based on their respective operation plans. A gas turbine engine suffers mechanical damage when it is turned on and operated, and it is assured to hold its required performance in a prespecified number of cumulative turning on and a certain cumulative operating period. So, the engine has to be overhauled before it exceeds the number of cumulative turnings on or the cumulative operating period, whichever occurs first. When a co-generation system is continuously operated throughout the year, the occasion to perform overhaul is strictly restricted, such as during the Christmas vacation period, because the overhaul needs a definite period and customers want to avoid the loss of operation.

We consider the following assumptions' policies:

- The j -th turning on and operation time of the system gives rise to an amount W_j of damage, where random variables W_j have an identical probability distribution $G(x)$ with finite mean, independent of the number of operation, where $\overline{G}(x) \equiv 1 - G(x)$. These damages are assumed to be accumulated to the current damage level. The cumulative damage $Z_j \equiv \sum_{i=1}^j W_i$ up to the j -th turning on and operation time has

$$\Pr\{Z_j \leq x\} = G^{(j)}(x) \quad j = 0, 1, 2, \dots, \tag{15.63}$$

where $Z_0 \equiv 0$, $G^{(0)}(x) \equiv 0$ for $x < 0$ and 1 for $x \geq 0$, and in general, $\Phi^{(j)}(x)$ is the j -fold Stieltjes convolution of $\Phi(x)$ with itself;

- When the cumulative damage exceeds a prespecified level K which the engine vendor prescribes, the customer of a co-generation system performs the engine overhaul immediately, because the assurance of engine performance expires otherwise. A cost c_K is needed for the sum of the engine overhaul cost and the intermittent loss of operation; and
- The customer performs the massive system maintenance annually, and checks all major items of the system precisely in several weeks. When the cumulative damage at such maintenance exceeds a managerial level k ($0 \leq k < K$) at which the customer prescribes, the customer performs the engine overhaul. A cost $c(z)$ is needed for the engine overhaul cost at the cumulative damage z ($k \leq z < K$). It is assumed that $c(0) > 0$ and $c(K) < c_K$, because it is not required to consider the loss of operational interruption.

15.5.2 Analysis

The probability that the cumulative damage is less than k at the j -th turning on and operation, and between k and K at the $j+1$ -th is

$$\int_0^k \left[\int_{k-u}^{K-u} dG(x) \right] dG^{(j)}(u). \quad (15.64)$$

The probability that the cumulative damage is less than k at the j -th turning on and operation, and more than K at the $j+1$ -th is

$$\int_0^k \bar{G}(K-u) dG^{(j)}(u). \quad (15.65)$$

It is evident that Equation 15.64 + Equation 15.65 = $G^{(j)}(k) - G^{(j+1)}(k)$.

When the cumulative damage is between k and K , the expected maintenance cost is, from Equation 15.64,

$$\sum_{j=0}^{\infty} \int_0^k \left[\int_{k-u}^{K-u} c(x+u) dG(x) \right] dG^{(j)}(u) = \int_0^k \left[\int_{k-u}^{K-u} c(x+u) dG(x) \right] dM(u). \quad (15.66)$$

where $M(x) \equiv \sum_{j=0}^{\infty} G^{(j)}(x)$. Similarly, when the cumulative damage is more than K , the expected maintenance cost is, from Equation 15.65,

$$c_K \int_0^k \bar{G}(K-u) dM(u). \quad (15.67)$$

Next, we define a random variable X_j as the time interval from the j -1-th to the j -th turning on and operation, and its distribution as $\Pr\{X_j \leq t\} \equiv F(t)$ ($j = 1, 2, \dots$) with finite mean $1/\lambda$. Then, the probability that the j -th turning on and operation occurs until time t is

$$Pr \left\{ \sum_{i=1}^j X_i \leq t \right\} = F^{(j)}(t). \quad (15.68)$$

From Equation 15.68, the mean time that the cumulative damage exceeds k at the j -th turning on and operation, is

$$\sum_{j=1}^{\infty} \int_0^{\infty} t [G^{(j-1)}(k) - G^{(j)}(k)] dF^{(j)}(t) = \frac{M(k)}{\lambda} \quad (15.69)$$

Therefore, the expected cost $C(k)$ per unit time is, from Ross (1983),

$$\frac{C(k)}{\lambda} = \frac{\int_0^k \left[\int_{k-u}^{K-u} c(x+u) dG(x) \right] dM(u) + c_K \int_0^k \bar{G}(K-u) dM(u)}{M(k)} \quad (15.70)$$

and the expected costs at $k=0$ and $k=K$ are, respectively,

$$\frac{C(0)}{\lambda} = \int_0^K c(x) dG(x) + c_K \bar{G}(K), \quad (15.71)$$

$$\frac{C(K)}{\lambda} = \frac{c_K}{M(K)} \quad (15.72)$$

15.5.3 Optimal Policy

We find an optimal damage level k^* which minimizes the expected cost $C(k)$ in Equation 15.70. Differentiating $C(k)$ with respect to k and setting it equal to zero,

$$[c_K - c(K)] \int_{K-k}^K M(K-x) g(x) dx + \int_0^K \left[\int_k^K g(x-u) dc(x) \right] M(u) du - c(k) = 0, \quad (15.73)$$

where $g(x) \equiv dG(x)/dx$ which is a density function of $G(x)$. When we denote the left-hand side of Equation 15.73 as $Q(k)$, we easily have

$$Q(0) = -c(0) < 0, \quad Q(K) = [c_K - c(K)]M(K) - c_K \quad (15.74)$$

Thus, if $Q(K) > 0$, i.e., $M(K) > c_K / [c_K - c(K)]$, then there exists a finite

k^* ($0 < k^* < K$) which minimizes $C(k)$, and the resulting cost is

$$\frac{C(k^*)}{\lambda} = \int_0^{K-k^*} [c(k^* + x) - c(k^*)] dG(x) + [c_K - c(k^*)] \bar{G}(K - k^*) \quad (15.75)$$

When $c(z) = c_1 z + c_0$ ($k \leq z < K$) where $c_1 K + c_0 < c_K$, Equations 15.70 and 15.71 are rearranged as, respectively,

$$\frac{C(K)}{\lambda} = \frac{\int_0^k \left[\int_{k-u}^{K-u} (c_1(u+x) + c_0) dG(x) \right] dM(u) + \int_0^k \bar{G}(K-u) dM(u)}{M(K)} \quad (15.76)$$

$$\frac{C(0)}{\lambda} = \int_0^K (c_1 x + c_0) dG(x) + c_K \bar{G}(K) \quad (15.77)$$

and $C(K)/\lambda$ is equal to Equation 15.72.

Differentiating $C(k)$ in Equation 5.76 with respect to k and putting it to zero, we have

$$(c_K - c_1 K - c_0) \int_{K-k}^K M(K-u) dG(u) - c_1 \int_{K-k}^K M(K-u) \bar{G}(u) du = c_0 \quad (15.78)$$

Letting $T(k)$ denote the left-hand side of Equation 15.78, we have

$$T(0) = 0, \quad T(K) = (c_K - c_1 K - c_0)[M(K) - 1] - c_1 K \quad (15.79)$$

Thus, if $T(K) > c_0$, i.e., $M(K) > c_K / (c_K - c_1 K - c_0)$, then there exists a finite k^* ($0 < k^* < K$) which minimizes $C(k)$.

Next, suppose that $G(x) = 1 - \exp(-\mu x)$, i.e., $M(x) = \mu x + 1$. Then, if $\mu K + 1 > c_K / (c_K - c_1 K - c_0)$, i.e., $\mu > (c_1 + c_0 / K) / (c_K - c_1 K - c_0)$, then there exists a finite k^* ($0 < k^* < K$). Further, differentiating $T(k)$ with respect to k ,

$$T'(k) = (\mu k + 1)e^{-\mu(K-k)}(c_K - c_1 K - c_0) \left(\mu - \frac{c_1}{c_K - c_1 K - c_0} \right) > 0 \quad (15.80)$$

since $(c_1 + c_0 / K) / (c_K - c_1 K - c_0) > c_1 / (c_K - c_1 K - c_0)$.

Therefore, we have the following optimal policy:

- If $\mu K > (c_1 K + c_0) / (c_K - c_1 K - c_0)$ then there exists a finite and unique k^* ($0 < k^* < K$) which satisfies

$$k e^{-\mu(K-k)} = \frac{c_0}{\mu(c_K - c_1 K - c_0) - c_1}, \quad (15.81)$$

and the resulting cost is

$$\frac{C(k^*)}{\lambda} = \frac{c_1}{\mu} \left(1 - e^{-\mu(K-k^*)} \right) + (c_K - c_1 K - c_0) e^{-\mu(K-k^*)} \quad (15.82)$$

- if $\mu K \leq (c_1 K + c_0) / (c_K - c_1 K - c_0)$ then $k^* = K$ and $C(K)/\lambda = c_K / (\mu K + 1)$.

15.5.4 Numerical Illustration

Suppose that $G(x) = 1 - \exp(-\mu x)$ and $c(z) = c_1 z + c_0$ ($k \leq z < K$). Then, the expected cost is, from Equation 15.76,

$$\frac{C(k)}{\lambda} = \frac{\frac{c_1[1 - e^{-\mu(K-k)}]}{\mu + c_1k + c_0} + (c_K - c_1K - c_0)}{\mu k + 1}$$

(15.83)

and the optimal policy is given in (15.81) and (15.82).

Table 15.6 gives the optimal managerial level k^* and its minimum cost $C(k^*)/\lambda$ when $c_1 = 0.1, 1$, $c_0 = 1, 10$, $c_K = 200, 1000$, $\mu = 0.5, 1$, and $K = 25, 50$. $C(k^*)$ s are smaller than $C(k)$ s and $C(k^*)/C(K)$ changes from 0.05 to 0.31 in this case. It is natural that k^* decreases when c_1, c_0 and $1/c_K$ decrease. The reduction of c_1 and c_0 ought to be equal to the increase of c_K . So, it is of interest in this illustration that $C(k^*)/\lambda$ decreases when c_1 and c_0 decrease, and $C(k^*)/\lambda$ slightly increases when c_K gains. It is obvious that k^* decreases and $C(k^*)/\lambda$ increases when K decreases. In this illustration, k^* decreases and $C(k^*)/\lambda$ increases when μ decreases.

The maintenance plan is settled at the beginning of co-generation system operation and the optimal managerial level is k^* calculated. The system is continuously operated and the cumulative damage is monitored. The system maintenance is performed annually and the customer decides whether the overhaul of the gas turbine engine should be performed or not by comparing the monitored cumulative damage and k^* .

Table 15.6. Optimal managerial level and k^* expected cost $C(k^*)/\lambda$

c_1	c_0	c_K	μ	K	k^*	$c(k^*)/\lambda$	$c(K)/\lambda$
1	1	200	1	50	41.3	1.02	3.92
0.1	1	200	1	50	41.0	0.12	3.92
1	10	200	1	50	43.6	1.23	3.92
1	1	1000	1	50	39.5	1.03	19.61
1	1	200	0.5	50	34.3	2.06	7.69
1	1	200	1	25	17.0	1.06	7.69

References

A-Hameed MS and Proschan F (1973) Nonstationary Shock Models. Stoch Process and Their Appl 1: 383–404
A-Hameed MS and Shimi IN (1978) Optimal Replacement of Damaged Devices. J of Appl Probab 15: 153–161
Barlow RE and Proschan F (1965) Mathematical Theory of Reliability. John Wiley & Sons, New York

- Bauer J, Cottrell DF, Gagnier TR and Kimball EW (1973) Dormancy and Power ON-OFF Cycling Effects on Electronic Equipment and Part Reliability, RADAC-TR-73-248 (AD-768 619)
- Boland PJ and Proschan F (1983) Optimal Replacement of a System Subject to Shocks. *Oper Res* 31: 697–704
- Brookner E (1985) Phased-array radars. *Sci Am* 252: 94–102
- Brookner E (1991) Practical Phased-Array Antenna Systems. Artech House, Boston
- Bucci OM, Capozzoli A and D'elia G (2000) Diagnosis of Array Faults from Far-Field Amplitude-Only Data. *IEEE Trans on Antennas and Propag* 48: 647–652
- Cahill MJ and Underwood FN (1987) Development of Digital Engine Control System for the Harrier. (Presented at AIAA/SAE/ASME/ASEE 23rd Joint Propulsion Conference, San Diego, California - June 29–July 2)
- Cottrell DF, Gagnier TR and Kimball EW *et al.* (1967) Dormant Operating and Storage Effects on Electronic Equipment and Part Reliability. RADAC-TR-67-307 (AD 821 988)
- Cottrell DF, Gagnier TR, Kimball EW and Kirejczyk TE, *et al.* (1974) Effects of Dormancy on Nonelectronic Components and Materials, RADAC-TR-74-269 (AD/A-002 838)
- Cox DR (1962) Renewal Theory, Methuen, London
- Davies WJ, Hoelzer CA and Vizzini RW (1983) F-14 Aircraft and Propulsion Control Integration Evaluation. *J of Eng for Power*, 105: 663–668
- Eccles ES, Simons ED and Evans JFO (1980) Redundancy Concepts in Full Authority Electronic Engine Control – Particularly Dual Redundancy, (Presented at AGARD Conference)
- Esary JD, Marshall AW and Proschan F (1973) Shock Models and Wear Processes. *The Annals of Probab* 1: 627–649
- Feldman RM (1976) Optimal Replacement with Semi-Markov Shock Models. *J of Appl Probab* 13: 108–117
- Feldman RM (1977) Optimal Replacement with Semi-Markov Shock Models Using Discounted Costs. *Math Oper Res* 2: 78–90
- Hesse JL (1975) Maintainability analysis and prototype operations. *Proceedings 1975 Annual Reliability and Maintainability Symposium*, pp.194–199
- Hevesh AH (1967) Maintainability of phased array radar systems. *IEEE Trans on Reliab R*-16: 61–66
- Ito K and Nakagawa T (1992) Optimal Inspection Policies for a System in Storage. *Comput Math Appl* 24: 87–90
- Ito K and Nakagawa T (1994) Optimal Inspection Policies for a System in Storage. *The J of Reliab Eng Associ of Jpn* 15(5): 3–7
- Ito K and Nakagawa T (1995) An Optimal Inspection Policy for a Storage System with High Reliability. *Microelectron and Reliab* 35(6): 875–886
- Ito K, Teramoto K and Nakagawa T (1999) Optimal Maintenance Policy for a Phased Array Radar. *The J of Reliab Eng Associ of Jpn* 21(4): 229–236
- Ito K and Nakagawa T (2003) Optimal Self-Diagnosis Policy for FADEC of Gas Turbine Engines. *Math and Comput Model* 38: 1243–1248
- Ito K and Nakagawa T (2004) Comparison of cyclic and delayed maintenances for a phased array radar. *J of the Oper Res Soci of Jpn* 47(1): 51–61
- Ito K and Nakagawa T (2006) Maintenance of a Cumulative Damage Model and Its Application to Gas Turbine Engine of Co-generation System. *Reliability Modeling, Analysis and Optimization, Series on Quality, Reliability and Engineering Statistics*, Vol.9 (Editor H. Pham), chapter 21, World Scientific Publishing Co. Pte. Ltd., pp.429–438
- Japan Ministry of Defense (2007) Defense of Japan (Annual White Paper)
- Keithley HM (1966) Maintainability impact on system design of a phased array radar. *Annual New York Conference on Electronic Reliability*, 7th, Vol.9, pp.1–10

- Kendell R (1981) Full-Authority Digital Electronic Controls for Civil Aircraft Engines. (Presented at Gas Turbine Conference and Products Show, Houston, Texas – March9-12)
- Luss H and Kander Z (1974) A preparedness model dealing with N-systems operating simultaneously. *Oper Res* 22(1): 117–128
- Malik DF and Mitchell JC (1978) Missile Material Reliability Prediction Handbook - Parts Count Prediction, (AD/A-053 403)
- Martinez EC (1984) Storage Reliability with Periodic Test. *Proceedings of Annual Reliability and Maintainability Symposium*, pp.181–185
- Menke JT (1983) Deterioration of electronics in storage. *Proceedings of National SAMPE Symposium*, pp. 966–972
- Nakagawa T (1980) Optimum inspection policy for a standby units. *J. of Oper Res Soc of Jpn* 23(1): 13–26
- Nakagawa T (1986) Modified discrete preventive maintenance policies. *Nav Res Logist Q* 33: 703–715
- Nakagawa T (2005) *Maintenance Theory of Reliability*, Springer-Verlag, London
- Nakagawa T (2007) *Shock and Damage Models in Reliability Theory*, Springer-Verlag, London
- Nakagawa T and Ito K (2007) Optimal Availability Models of a Phased Array Radar. *Recent Advances in Stochastic Operations Research*, World Scientific Publishing Co. Pte. Ltd., pp.115–130
- Posner MJM and Zuckerman D (1984) A Replacement Model for an Additive Damage Model with Restoration. *Oper Res Lett* 3: 141–148
- Puri PS and Singh H (1986) Optimum Replacemnet of a System Subject to Shocks: a Mathematical Lemma. *Oper Res* 34: 782–789
- Robinson K (1987) Digital Controls for Gas Turbine Engines. (Presented at the Gas Turbine Conference and Exhibition, Anaheim, California – May31–June4)
- Ross SM (1970) *Applied probability models with optimization application*, Holden-Day, San Francisco
- Ross SM (1983) *Stochastic Processes*, John Wiley & Sons, New York
- Scoles RJ (1986) FADEC – Every Jet Engine Should Have One, (Presented at Aerospace. Technology Conference and Exposition, Long Beach, California – October 13–16)
- Shima E and Nakagawa T (1984) Optimum inspection policy for a protective device. *Reliab Eng* 7(2): 123–132
- Skolnik MI (1980) *Introduction to Radar Systems*, McGraw-Hill Publishing Company, New York
- Skolnik MI (1990) *Radar Handbook*, McGraw-Hill Publishing Company, New York
- Smith HB, Jr. and Rhodes C, Jr. (1982) Storage Reliability of Missile Materiel Program - Storage Reliability Prediction Handbook for Part Count Prediction, (AD/A-122 439)
- Taylor HM (1975) Optimal Replacement Under Additive Damage and Other Failure Models. *Nav Res Logist Q* 22: 1–18
- Thomas LC, Jacobs PA and Gaver DP (1987) Optimal inspection policies for standby systems. *Stoch Model* 3(2): 259–273
- Trapp RD, Graber RD, Farmer WD and Luhrs RA (1981) An Approach for Assessing Missile System Dormant Reliability, BDM/A-81-016-TR (AD/A-107 519)
- U.S. Congress, Office of Technology Assessment (1992) *After the Cold War: Living With Lower Defense Spending*, OTA-ITE-524
- Witte LC, Schmidt PS and Brown DR (1988) *Industrial Energy Management and Utilization*, Hemisphere Publishing Corporation, New York
- Zacks S and Fenske WJ (1973) Sequential determination of inspection epochs for reliability systems with general lifetime dustributions. *Nav Res Logist Q* 20(3): 377–386

Zuckerman D (1977) Replacement Models Under Additive Damage. Nav Res Logist Q 24: 549–558

Zuckerman D (1980) A Note on the Optimal Replacement Time of Damaged Devices. Nav Res Logist Q 27: 521–524

Reliability Centered Maintenance

Atiq Waliullah Siddiqui and Mohamed Ben-Daya

16.1 Introduction

The maintenance function must ensure that all production and manufacturing systems are operating safely and reliably and provide the necessary support for the production function. Furthermore, maintenance needs to achieve its mission using a cost-effective maintenance strategy. What constitutes a cost effective strategy evolved over time?

In the past, it was believed every component of a complex system has a right age at which complete overhaul is needed to ensure safety and optimum operating conditions. This was the basis for scheduled maintenance programs. The limitation of this thinking became clear when it was used to develop the preventive maintenance program for the “new” Boeing 747 in the 1960s. The airlines knew that such a program would not be economically viable and launched a major study to validate the failure characteristics of aircraft components. The study resulted in what became the Handbook for the Maintenance Evaluation and Program Development for the Boeing 747, more commonly known as MSG-1 (Maintenance Steering Group 1). MSG-1 was subsequently improved and became MSG-2 and was used for the certification of DC 10 and L 1011. In 1979 the Air Transport Association (ATA) reviewed MSG-2 to incorporate further developments in preventive maintenance; this resulted in MSG-3, the Airline/Manufacturers Maintenance Program Planning Document applied subsequently to Boeing 757 and Boeing 767.

United Airlines was sponsored by the US Department of Defense to write a comprehensive document on the relationships between Maintenance, Reliability and Safety. The report was prepared by Stanley Nowlan and Howard Heap (Nowlan and Heap 1978) it was called ‘Reliability Centered Maintenance’. The application of MSG-3 outside the aerospace industry is generally known as RCM. Afterwards, RCM spread to nuclear power plants and other industries.

The studies in the airline industry revealed that scheduled overhaul did not have much impact on the overall reliability of a complex item unless there is a dominant failure mode. Also, there are many items for which there is no effective form of

scheduled maintenance. In Figure 16.1, it is clear that only 11% of the components exhibit a failure characteristic that justify a scheduled overhaul or replacement. Eighty nine percent showed random failure characteristics for which a scheduled overhaul or replacement was not effective. Therefore, new thinking is required to deal with the remaining 89%.

These findings redefined maintenance by focusing thinking on system function rather than operation. To understand better this shift in thinking and introduce a formal definition of RCM, the Society of Automotive Engineers has developed and issued SAE JA-1011, which provides some degree of standardization for the RCM process. The SAE standard defines the RCM process as asking seven basic questions from which a comprehensive maintenance approach can be defined:

1. What are the functions and associated performance standards of the asset in its present operating context?
2. In what ways can it fail to fulfill its functions?
3. What causes each functional failure?
4. What happens when each failure occurs?
5. In what way does each failure matter?
6. What can be done to predict or prevent each failure? and
7. What should be done if a suitable proactive task cannot be found?

From these seven questions emerges a systematic process to determine the maintenance requirements of any physical asset in its operating context, called Reliability Centered Maintenance. The first step in the RCM process is to define the functions of each asset in its operating context, together with the associated desired standards of performance. Then identify what failure can occur and defeat the functions. Once each functional failure has been identified, the next step is to try to identify the causes of failures, *i.e.*, all the events which are reasonably likely to cause each failure mode. These events are known as failure modes. The fourth step in the RCM process involves listing failure effects, which describe what happens when each failure mode occurs at the local and system level. The RCM process classifies these consequences into four groups, as follows:

- Hidden failure consequences;
- Safety and environmental consequences;
- Operational consequences; and
- Non-operational consequences.

Identifying the consequences of failure helps in prioritizing the failure modes because failures are not created equal. By now the RCM process generated a wealth of information on how the system works, how it can fail, and the causes and consequences of failures. The last step is to select maintenance tasks to prevent or detect the onset of failure. Only applicable and effective tasks are selected.

This way RCM can be used to create a cost-effective maintenance strategy to address dominant causes of equipment failure. It is a systematic approach to defining a preventive maintenance program composed of cost-effective tasks that preserve important systems functions.

The RCM framework combines various maintenance strategies including time-directed preventive maintenance, condition based maintenance, run-to-failure, and proactive maintenance techniques in an integrated manner to increase the probability that a system or component will function in the required manner in its operating context over its design life-cycle. The goal of the method is to provide the required reliability and availability at the lowest cost. RCM requires that maintenance decisions be based on clear maintenance requirements that can be supported by sound technical and economic justification.

The purpose of this chapter is to provide an informative introduction to RCM methodology and is organized as follows: in the next section, RCM philosophy along with its principles, key features, goals and benefits are discussed. This is followed by discussion on background issues, including system, system boundary, interfaces and interactions. Section 3 talks about failure and its nature. Section 4 presents RCM methodology and practical RCM Implementation issues are discussed in Section 5. The last section concludes the chapter.

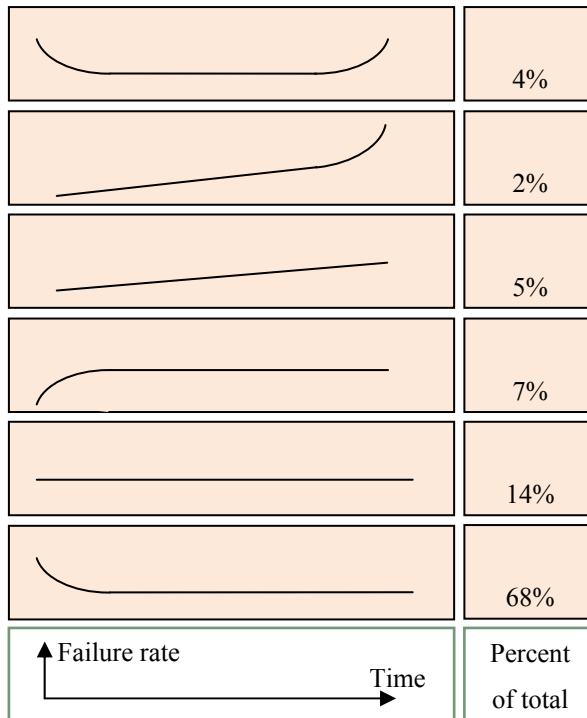


Figure 16.1. Aircraft failure characteristics (Nowlan and Heap, 1978)

16.2 RCM Philosophy

Reliability Centered Maintenance philosophy is based on a system enhancement method that keeps a cost effective view while identifying and devising operational, and maintenance policies and strategies. This is done in order to manage the risks of a system's functional failure in an economically effective manner, and is especially applicable to situations where there are low or constrained financial resources.

RCM philosophy fundamentally differs from other maintenance strategies, by preserving system functionality to a desired level, as opposed to maintaining equipments keeping it isolated with their relationship to the system. In summary, Reliability Centered Maintenance is a systematic approach to defining a planned maintenance program poised of cost-effective tasks while preserving critical plant functions.

An important aspect of this philosophy is to prioritize systems by assigning levels of criticality based on the consequences of failure. This aspect, in particular, is in line with the fundamental objective of being cost effectiveness with efficiency channelizing the resources to the high priority tasks. This is done by identifying required design and operational modifications and *justified* maintenance strategies according to the priority levels. As an example, equipment that is non-critical to the plant may be left to *run to failure* while equipment serving critical functions is preserved at all cost. Maintenance tasks are selected to address the dominant failure causes addressing preventable failures through maintenance. RCM underlines the use of predictive maintenance (PdM) besides traditional preventive measures.

16.2.1 RCM Principles and Key Features

There are four principles or key features that characterize the RCM process. These features are:

1. *Preserving the system function* is the first and principal feature of RCM process. This feature is important in its understanding. It must be stressed, as it forces a change in the typical view of equipment maintenance and replaces it with the view of functional preservation. What is required is to identify the desired system output and ensure availability of the same output level?
2. *Identification of the particular failure modes that can potentially cause functional failure* is the second feature of RCM process. This information is crucial whether a design or operational modification is required or a maintenance plan is to be made.
3. *Prioritizing key functional failures* is the third of the RCM process features. This feature is of foremost importance as the philosophy of efficiency with cost effectiveness can be achieved through this feature. Efforts and resources are dedicated to equipment supporting critical functions and their unavailability means major degradation of plant to even total shutdown.
4. *Selection of applicable and effective maintenance tasks for the high priority items* is the fourth feature of the RCM process. As described earlier, the

purpose of prioritizing is to make an efficient and cost effective use of resources.

16.2.2 RCM Goals and Benefits

Various goals are served through RCM implementation. First, it helps determine the optimum maintenance program. It is also a proven and effective strategy in optimizing the maintenance efforts, both in terms of, operational efficiency and cost effectiveness. It helps keeping focal point on maintaining or preserving the most crucial system functions, while averting maintenance actions that are not particularly required. In essence it endeavors for the required system reliability at the lowest possible cost without forgoing issues related to the safety and the environment.

Significant benefits are also tangible; these typically includes cost saving, shifting from time-based to condition-based work, spare parts usage reduction, improved safety and environmental conditions, improvement in workload reduction and operation performance, large information database enhancing the level of skill and technical knowledge.

16.2.3 System, System Boundary, Interfaces and Interactions

A better understanding of RCM methodology requires understanding of few key systems definitions. This section briefly discusses such key terms.

16.2.3.1 Systems

All systems are made up of three basic components. These are *input*, *process* and *output*. This is shown in Figure 16.2. The figure shown is also known as a *basic system diagram* which is one of the ways to *model* or represent any system.

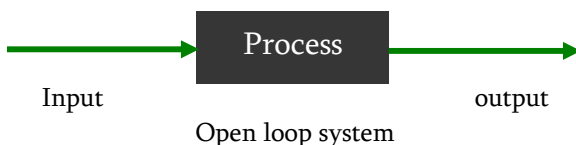


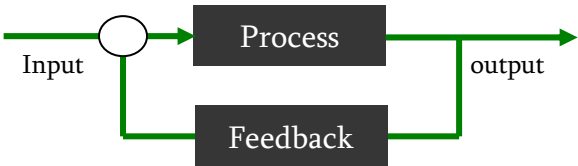
Figure 16.2. Basic system diagram of an open loop system

The model of a system shown in Figure 16.2 is also known as an *open loop system*. An open loop system is defined as a system that has no feedback. As opposed to an open loop system, a *closed loop system* (Figure 16.3) uses a *feedback* to measure the output ensuring actual results seeking desired results.

16.2.3.2 Complex Systems

Industrial systems are almost always *complex* in nature. The term *complex systems* refers to a system in which the elements are *varied* and have *complex or convoluted* relationships with other elements of the system. The systems which are not complex in nature generally involve fewer engineering disciplines, *e.g.*, a

washing machine is an electro-mechanical system. Examples of some complex technological systems, signifying the three basic components, are illustrated in Table 16.1.



Closed loop systems

Figure 16.3. Basic system diagram of a closed loop system

Table 16.1. Examples of complex system

System	Inputs	Process	Outputs
Weather satellite	Images, signals	Data storage, processing and transmission	Processed images
Airlines ticketing system	Travel requests	Data management	Reservation and air tickets
Oil refinery	Crude oil, catalysts, energy	Cracking, separating and blending	Petrol, diesel and lubricants <i>etc.</i>
Nuclear power plant	Fuel (uranium), heavy water	Fission reaction, power generation	Electric a.c. power
road cargo system	cargo request	map tracing, communication	Routing information, cargo delivery

16.2.3.3 Modeling a Complex System

By character, complex systems can be made up of a number of major *systems* which are composed of further more simple working elements down to *primitive elements* such as gears, pulleys, buttons, resistors, and capacitors, *etc.*

The architecture, also known as system block diagram (see Figure 16.4.), shows the structure and terminologies used to model a complex system (*here a complex systems means a plant or a facility*). As can be seen, the highest level is known as a *plant* having the largest scope. This is followed by a number of systems with smaller scopes. Collection of all these systems makes a plant or complex system. Each system is made up of *components*, and each component has a simpler functionality as compared to a systems. These components are the first to provide a significant functionality. For this reason, the components are considered to be the basic system building blocks.

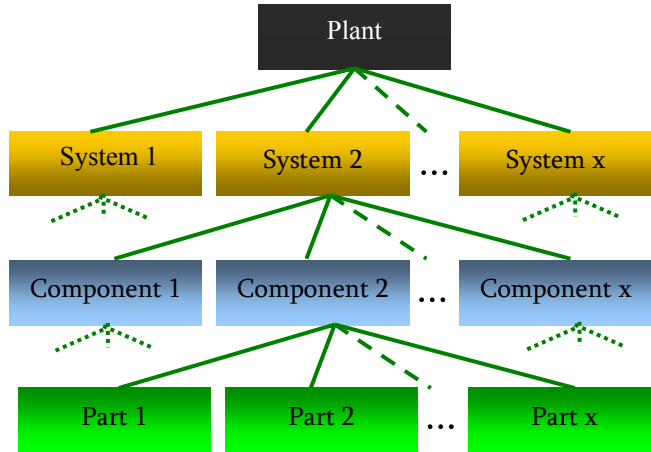


Figure 16.4. Architecture of a complex system

The main purpose of a system is to alter the three basic entities on which a system, generally, operates. These are *information, material and energy*, which provide us a good basis to classify principal functional elements. These are:

1. Signal (a system can generate, transmit, distribute and receive signals used in sensing and communication);
2. Data (a system can analyze, organize, interpret, or convert data into forms that a user desires);
3. Material (provide structural support for a system– it can transform shape or composition of materials, *etc.*); and
4. Energy (provide energy to a system).

Components are defined as physical embodiment of these functional elements which can be classified in six groups as shown in Figure 16.5. These six categories are electronic, mechanical, electromechanical, thermo-mechanical, electro-optical, and software.

The lowest or the most primal level in a system is known as parts. A part in itself does not have any functioning but are required to put together components. Examples of parts are: electronic: LED, resistors, transistors; mechanical: gears, ropes, pulleys, seals; electromechanical: wires, couplings, magnets; thermo-mechanical: coils, valves; electro-optical: lenses, mirrors; software: algorithms *etc.*

Interfaces and interactions

There are three types of interface that may occur in a system. These are:

1. *Connectors*: connectors facilitate the transmission of physical interaction, *e.g.*, transmission of fluid through pipes or electricity through cables, *etc.*;

- 2. *Isolators*: isolators impede or block physical interaction, *e.g.*, rubber cover over copper wire, *etc.*; and
- 3. *Converters*: converters alter the form of the physical medium, *e.g.*, pump changes the force in a fluid, *etc.*

More examples of interfaces along with type of physical medium is given in Table 16.2.

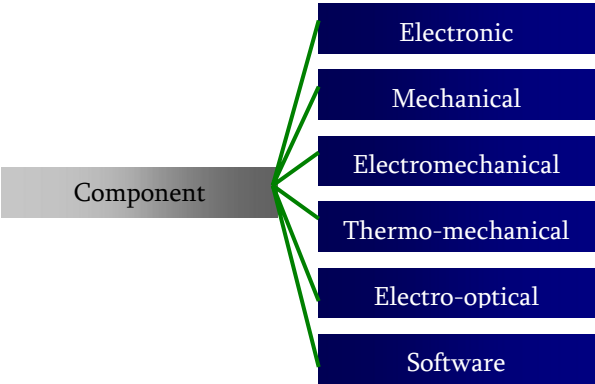


Figure 16.5. Classification of component

Table 16.2. Examples of various types of interfaces

Type (medium)	Electrical (current)	Mechanical (force)	Hydraulic (fluid)	Human- Machine (information)
Connectors	Cable, switches	Cam shaft, connecting rod	Value, piping	Control display panel
Isolators	Insulator	Bearing, shock absorbers	Hydraulic Seal	Window shield
Converters	Transformer, antenna	Crank shaft, gear train	Pump, nozzle	Software

16.3 Failure and its Nature

Understanding of failure and its nature is at the core of understanding and implementing the RCM strategy. A look at this aspect is required before moving forward. Key definitions are presented below.

Failure

Failure of a component occurs when there is a significant deviation from its original condition that renders it unacceptable for its user. It can be categorized as

complete failure, partial failure, intermittent failure, failure over time, or over-performance of function.

Functional failure

Functional failure on the other hand is defined as the inability of a system to meet its specified performance standard.

Potential functional failure

This is an identifiable physical condition that identifies an impending functional failure.

Failure modes

Failure modes are defined as the manner in which a failure may happen. It could be physical such as conditions where a part fails or conceptual where failure is not identified and organizational where absence of well defined job roles and mission priorities leads to failures.

Reliability

Reliability is the ability of a system or a component to perform its required functions consistently under the stated conditions for a specified period of time or in other words it is the capacity of a device or system to resist failure.

16.4 RCM Methodology

RCM has a seven step methodology. This methodology warrants documentation that records exactly how maintenance tasks were selected and why these were the best possible selections amongst a number of competing alternatives. These seven steps include:

1. Selecting systems and collecting information;
2. System boundary definition;
3. System description and functional block diagram;
4. System functions and functional failure;
5. Failure mode and effective analysis (FEMA);
6. Logic decision tree analysis (LTA); and
7. Task selection.

The next sections describe each of these steps.

16.4.1 Selecting Systems Selection and Collecting Information

As discussed earlier, and by experience, system level analysis is the best approach; component level lacks defining significance of functions and functional failure, while plant level analysis makes the whole analysis readily intractable.

Having decided that system is the best practical level for conducting such an analysis, the next question to confront is to choose what systems and in which order. One answer could be to select all systems within the plant or facility and in any order. However, this contradicts the spirit and the main drive of RCM – cost effectiveness. This argument is also supported by the fact that many systems neither have the history of consistent failures nor incur excessive maintenance costs that would justify the whole effort. As this may be a situation faced at most plants several selections schemes, that are employed, can be identified as follows:

1. Systems with a large number of corrective maintenance tasks during recent years;
2. Systems with a large number of preventive maintenance tasks and or costs during recent years;
3. A combination of scheme 1 and 2;
4. System with a high cost of maintenance of corrective maintenance tasks during recent years;
5. Systems contributing significantly towards plant outages/shutdowns (full or partial) during recent years;
6. Systems with high concern relating to safety; and
7. Systems with high concern relating to environment.

It has been found with experience that all of these schemes except schemes 6 and 7 yield more or less the same results. What is a suitable scheme in a particular case is a subjective matter, but more importantly it should be done in as simplistic a way as possible with a minimal expenditure of time and resources. An indicator of decent selection is that systems chosen for an RCM program are easily pinpointed without a big margin of error.

The next step, after selecting systems, is collecting information related to these systems. A good practice is to start collecting key information and document right at the outset of the process. Some common documents are identified that may be required in a typical RCM study. These are:

- P&ID (piping and instrumentation) diagram.
- systems schematic and/or block diagram (usually less messy than P&ID and facilitates better understanding of main equipment).
- Functional flow diagram (usually less messy than P&ID and facilitates better understanding of functional features of the system).
- Equipment design specification and operations manuals (a source of finding design specifications and operating condition details).
- Equipment history (failure and maintenance history in specific).
- Other identified sources of information, unique to the plant or organizational structure. Examples include industry data for similar systems.
- Current maintenance program used with the system. This information is generally not recommended to collect before step 7, in order to avoid and preclusions and biases that may affect the RCM process.

16.4.2 System Boundary Definition

The identification of a system depends on various factors. These may include plant complexity, governmental or regulatory rules and constraints, local and / or unique industry practices, a firm's financial structure, *etc.* Although a gross system's definitions and boundaries have been identified for specific cases, that may be used to good effect in step one as well but does not suffice for further analysis. Detailed and precise boundary identification is vital. Key reasons for this are:

1. An exact knowledge of what is included (conversely not included) in a system in order to make sure that any key system function or equipment is not neglected (conversely not overlapped from another equipment). This is especially important if two adjacent systems are selected.
2. Boundary definition also includes system interfaces (both IN and OUT interfaces) and interactions that establish inputs and outputs of a system. An accurate definitions of IN and OUT interfaces is a precondition to fulfil step 3 and 4.

There are no clear rules to define system boundaries; however as a general guideline a system has one or two main functions with a few supporting functions that would make up a logical grouping of equipment. However the boundary is identified, there must be clear documentation as part of a successful process.

16.4.3 System Description and Functional Block Diagram

The logical step to follow after system selection and boundary definition is to analyze further and document the necessary details of the systems under scope. This step generally involves form to document baseline characterization of a system that is eventually to be used in stipulating PM tasks. A typical form is shown in Figure 16.6.

The five items established during this step are as follows:

1. *System description*

In this step data already collected in earlier stages are put in the system analysis form. An accurate and well documented system definition will help produce concrete payback. This baseline information also serves as a record that will assist in comparisons during modifications and upgrades in the design or operations. It also identifies key design and operational parameters that directly affect the performance of the system functions.

2. *Functional block diagram*

A system block diagram, as discussed previously (Figure 16.4.), deals with the static and physical relationship that exists in a system. It does not illustrate the more significant characteristics of a system such as the behavioral response that happens with the changes in the system environment. This behavioral response depends on the function that a system can perform to such environmental inputs and constrictions. To model this functional behavior a FBD or functional block

diagram (Figure 16.7) is used. FBD elaborates functional flow in a system which is a top-level representation of the major function that a system performs. Arrows connecting blocks roughly represent interaction amongst functions and with the IN/OUT interfaces (to be adjoined in the next step).

RCM System Analysis (system description)		
Date:	Plant:	Location:
System Name:	RCM Analyst(s): 1. 2.	
System ID:		
System Location:		
Functional Description		
Key Parameters		
Key equipment		
Redundancy Features		
Safety Features		

Figure 16.6. Typical RCM system analysis form

Figure 16.7 shows example of functional flow block diagram for a car temperature control system. The system has two main functions: temperature detection and cooling control. Each function is further explicated by FBD discretely.

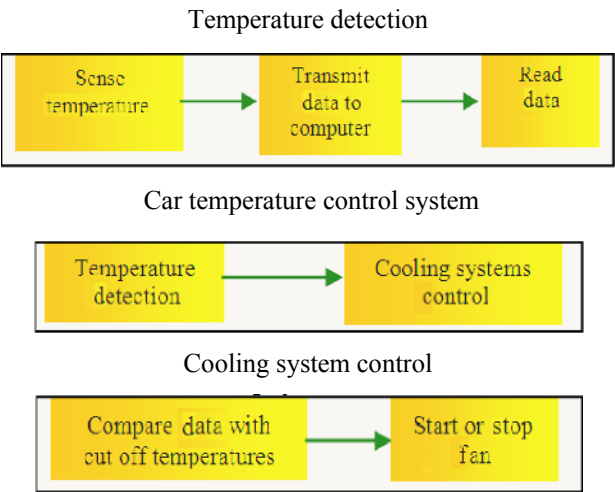


Figure 16.7. Functional flow block diagram

3. In/out interfaces

After defining a system along with its boundary and major function we can define system interfaces. IN interfaces exist within a system while OUT interfaces exist at

the boundaries of the system, making themselves the principle objects to preserve system functions. A point to note is that the IN interfaces might be OUT interfaces in some other systems. If an interface is within a system boundary connecting to system environment it is called the Internal OUT interface. Likewise, in step 3 a form is used to document interfaces (see Figure 16.8).

RCM System Analysis (interface definition)		
Date:	Plant:	Location:
System Name:	RCM Analyst(s):	
System ID:	1.	
System Location:	2.	
IN interfaces		
OUT interfaces		
Internal OUT interfaces		

Figure 16.8. Typical RCM system analysis form for interface definition

4. *Systems work breakdown structure*

Systems work breakdown structure or SWBS is a term used to identify a list of equipment/components for each of the function shown in a functional block diagram. This list is defined at the component level of assembly that resides with the system boundary. Identification of all components within a system is essential as otherwise it will eliminate these unlisted components out of the PM considerations. A typical SEBS form is shown in Figure 16.9.

RCM System Analysis (System Work Breakdown Structure)		
Date:	Plant:	Location:
System Name:	RCM Analyst(s):	
System ID:	1.	
System Location:	2.	
	Item	Number of item used
Non-instrumentation List		
Instrumentation List		

Figure 16.9. Typical RCM system analysis form for system work breakdown structure

5. *Equipment history*

Equipment history is also recorded in a form as shown in Figure 16.10. It contains failure history that has been experienced during the last couple of years. This data can be obtained from work orders used for corrective and preventive maintenance.

RCM System Analysis (Equipment History)			
Date:	Plant:		Location:
System Name: System ID: System Location:			RCM Analyst(s): 1. 2.
Component	Date	Failure Mode	Failure Cause

Figure 16.10. Typical RCM system analysis form for equipment history

16.4.4 System Functions and Functional Failure

This step identifies the functions that are needed to be preserved by the system (at the OUT interfaces). An important point to note is that these statements are for defining system functions and not the equipment. With the definition of system functions comes the functional failures. In fact, failing to preserve a system function constitutes what is called a functional failure. This leads to the step of how a process function can be defeated. This requires two things; keeping the focus on the loss of function and not the equipment and that the functional failures are more than just a single statement of loss of function. The loss conditions may be two or more (e.g., complete paralysis of the plant or major or minor deprivation of functionality. This distinction is important and will lead to the proper ranking of functions and functional failures.

16.4.5 Failure Mode and Effective Analysis (FEMA)

Failure Modes and Effects Analysis (FMEA) is a fundamental tool used in reliability engineering. It is a systematic failure analysis technique that is used to identify the failure modes, their causes and consequently their fallouts on the system function.

As discussed in Chapter 4, identifying known and potential failure modes is an important task in FMEA. Using data and knowledge of the process or equipment, each potential failure mode and effect is rated in each of the following three factors:

- *Severity* – the consequence of the failure when it happens;
- *Occurrence* – the probability or frequency of the failure occurring; and
- *Detection* – the probability of the failure being detected before the impact of the effect is realized.

Then these three factors are combined in one number called the risk priority number (RPN) to reflect the priority of the failure modes identified. The risk priority number (RPN) is simply calculated by multiplying the severity rating, times the occurrence probability rating, times the detection probability rating.

FMEA process is usually documented using a matrix similar to the one shown in Figures 4.1–4.3 (see Chapter 4 for more details). For each component, the failure modes are listed, their causes are identified, and their effects are determined. This initial screening of the failure modes help to prioritize them. Further prioritization will be conducted in the next step using logic tree analysis.

16.4.6 Logic or Decision Tree Analysis (LTA)

Logic tree or decision tree analysis (LTA) is the sixth step in RCM methodology. The purpose of this step is to prioritize further the resources that are to be committed to each failure mode. This is done since each failure mode and its impact on the whole plant is not the same. Any logical scheme can be adopted to do this ranking. RCM processes a simple and intuitive three question logic of decision structure that enables a user, with minimal effort, to place each failure mode into one of the four categories. Each question is answered yes or no only. Each category which is also known as bin forms natural segregation of items of respective importance. The LTA scheme is shown in Figure 16.11.

This makes items fall in the categories of A, B, C, D/A, D/B or D/C. For the priority scheme, A and B have higher priority over C when it come to allocation of scarce resources and A is given higher priority than B. In summary, the priority for PM task goes in the following order:

- A or D/A;
- B or D/B; and
- C or D/C.

16.4.7 Task Selection

In this step, we have to allocate PM tasks and resources and this is the point where we would be able to reap the maximum economic benefits of RCM activity. The task selection requires that each task is applicable and effective. Here, applicable means that the task should be able to prevent failures, detect failures, or unearth hidden failures, while effective is related to the cost effectiveness of the alternative PM strategies. If no PM task is selected the only option is to run equipment to failure. This activity requires contribution from the maintenance personnel as their experience is invaluable in the right selection of the PM task.

16.5 RCM Implementation

The practical side or implementation of RCM is an important factor to look at since; typically, such a program will initially focus on its planning and completion of systems analysis phase. However, in reality the real complexity that is almost always present in planning and coordination phases of such efforts is hard to realize before it catches up. This immediately results in delays and problems in communication, decision and consequently in the execution of the project. In this

case certain practical guidelines present in RCM reference have been summarized below.

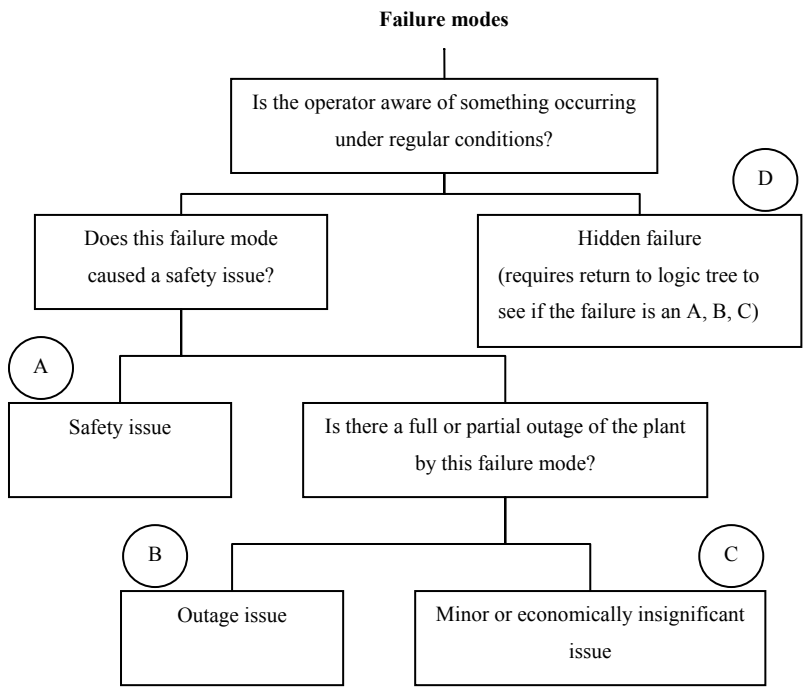


Figure 16.11. Logic tree analysis

16.5.1 Organizational Factors

Organizational factors play an important part as they define responsibilities and jurisdictions, and establish communication channels essential for such an effort. Issues that are needed to be addressed are:

1. Company organization: although, a prime factor in the success of any such concerted and complex effort requires motivation and strong personalities, a loose organizational structure with unclear responsibilities' and loose communication channels proves to be a major hurdle in making any such effort a success. With goo team work, where there is top management commitment and clear organizational hierarchy, the already complex task is not further aggravated, and this setup rather supports and simplifies the project handling. In RCM a key success factor is the separation of production and maintenance functions with clear peer like coordination.
2. The decision making process: any large scale and complex successful projects is driven by commitment from the management, both at the top and at plant level. This becomes even more essential as such initiative demands a change

in a company's cultural and major operational methodologies, resulting in phenomena such as internal resistances and lack of employee commitment and motivation. This may prove to be a real threat as resulting quality compromise renders the whole process futile.

3. The financial aspect: financial commitments are required while unforeseen costs may appear. Major cost factors include training, consultation fees, software support, project facilitation cost, *etc.*
4. Project ownership (The Buy-in Factor): buy-in process signifies a process where individuals or teams responsible for implementation are made part of the planning and development process, creating a sense of ownership. This proves to be a motivating factor which contributes towards removing project hurdles and success.

16.5.2 RCM Teams

RCM team formation is another issue that is almost always present in RCM projects. Availability of experienced personnel and on-site plant staff with the present work load are some of the issues to handle, especially for keeping the buy-in factor in view. Various resource allocation strategies are mentioned in the literature; one good strategy is to assign appropriate on site personnel to the RCM team by giving it top priority over other activities. Another strategy is to increase plant staffing if current staffing is not committed. A third strategy is to commit a team from corporate head quarters and a fourth strategy is to outsource or contract the RCM project.

As for the team formation, a typical team comprises four to five members with a facilitator. Diverse experience proves healthy for the team. The facilitator in the team is generally responsible for the coordination of efforts and guides in achieving buy-in during the early stages of the projects.

16.5.3 Scheduling Consideration and Training

Scheduling considerations also play a key role in RCM success. Lack of dedicated allocation of team personnel and other resources severely hinders project deadlines and this is common in such situations. The scheduling considerations not only involve project management aspects and logistics but they must include a timeframe that would enable the organization to pass through the learning curve that is needed in the change in the mindset and culture. The schedule must also include a pilot project.

Training is also required with a firm grip on RCM philosophy, seven step methodology and; a good knowledge of practical issues and understanding of the current maintenance situation, *etc.* Generally, training for RCM is carried out in a two step method form. In the initial step, a classroom setting for training works well, with a session of 3–5 days, while hands-on training cannot be avoided due to the nature of the process. This hands-on training is best done under the guidance of a trained person generally available as facilitator of the project. The important aspects to keep in mind include how to get acquaintance and documentation of the current maintenance situation, the knowledge of RCM, its methodology, and in

what ways RCM would help the plant in terms of cost effectiveness and plant efficiency.

16.6 Conclusion

A brief description of RCM is presented in this chapter. The objective was to introduce the reader to the basics of RCM. The methodology has proved time and again to deliver fruitful results. However, in spite of the simplistic and intuitive appeal, application without full understanding may lead to project hiccups if not total failures. A detailed pre-study is required before such an initiative should be undertaken.

One should be careful that the initial simplistic appeal of the methodology should not make a user unsighted to the real application issues and challenges. A lack of experience as RCM implementers and/or people providing necessary information may hinder in the success of the project. Management's direct interest is always crucial and any such activity should not be undertaken until or unless there is full support, commitment and involvement from both top and plant management. Buy-in is a factor that should never be forgotten. With cultural and fundamental work methods changes at hand, buy-in is a proven strategy to confront internal and cultural resistances. With a learning curve required to grasp fully the philosophy of method, initial investments on training also serves well.

RCM is a highly intuitive and applicable method. Its philosophy and methodology was discussed along with implementation issues and challenges. Deciding to use this methodology with a good handling of implementation challenges ensures considerable efficiency and economic benefits.

References

- Anderson RT, Neri L (1990) Reliability Centered Maintenance: Management and Engineering Methods, Elsevier Applied Sciences, London.
- August J (1999) Applied reliability-centered maintenance. Oklahoma: Penn Well.
- Blanchard BS (1992) Logistics Engineering and Management. Prentice-Hall, Englewood Cliffs, NJ.
- http://en.wikipedia.org/wiki/Failure_mode_and_effects_analysis
- http://en.wikipedia.org/wiki/Reliability_%28statistics%29
- http://en.wikipedia.org/wiki/Reliability_centered_maintenance
- http://en.wikipedia.org/wiki/Reliability_engineering
- http://safety.eng.umanitoba.ca/downloads/fmea_presentation.ppt
- https://www.fs.com/customer/MFG_Best_Practices_Session_5.ppt
- Kelly A (1984), Maintenance Planning and Control, Butterworths, London.
- Lenaerts H (1998) Optimalisatie van het Onderhoud m.b.v. Onderhoudsconcepten. KULeuven, Master thesis.
- Moubray J (1997) Reliability Centered Maintenance. Butterworth-Heineman, Oxford.

Nowlan FS, Heap HF (1978) Reliability Centered Maintenance, United Airlines Publications, San Francisco, CA.

Smith AM and Hinchcliffe GR (2004) RCM gateway to world class maintenance. Burlington : Elsevier Butterworth-Heinemann.

Total Productive Maintenance

P.S. Ahuja

17.1 Introduction to TPM

Manufacturing organizations worldwide are facing many challenges to achieve successful operation in today's competitive environment. Modern manufacturing requires that, to be successful, organizations must be supported by both effective and efficient maintenance practices and procedures. The global marketplace has necessitated many organizations to implement proactive lean manufacturing programs and organizational structures to enhance their competitiveness (Bonavia and Marin, 2006). Over the past two decades, manufacturing organizations have used different approaches to improve maintenance effectiveness. One approach to improving the performance of maintenance activities is to develop and implement strategic TPM programs (Ahuja and Khamba, 2007). Among various manufacturing programs, Total Quality Management (TQM), Just-in-Time (JIT), Total Productive Maintenance (TPM) and Total Employee Involvement (TEI) programs have often been referred to as components of "World Class Manufacturing" (Cua *et al.* 2001).

According to Nakajima (1988), vice-chairman of Japan Institute of Plant Maintenance, TPM is a combination of American preventive maintenance and Japanese concepts of total quality management and total employee involvement. TPM is a methodology originated by Japan to support its lean manufacturing system. TPM is a proven manufacturing strategy that has been successfully employed globally for achieving the organizational objectives of core competence in the competitive environment. TPM implementation methodology provides organizations with guidelines to transform fundamentally their shop-floor by integrating culture, process and technology.

Total Productive Maintenance (TPM) as the name suggests consists of three words:

<i>Total:</i>	signifies to consider every aspect and involving everybody from top to bottom;
---------------	--

- Productive:* emphasis on trying to do it while production goes on and minimize troubles for production; and
- Maintenance:* means equipment upkeep autonomously by production operators in good condition – repair, clean, grease, and accept to spend necessary time on it.

TPM is considered to be Japan's answer to U.S. style productive maintenance. TPM is a Japanese concept developed in the 1970s by extending preventive maintenance to become more like productive maintenance. TPM is an innovative approach to plant maintenance that is complementary with TQM, JIT, TEI, Continuous Performance Improvement (CPI), and other world-class strategies (Cua *et al.* 2006). TPM has been widely recognized as a strategic weapon for improving manufacturing performance by enhancing the effectiveness of production facilities. Originally introduced as a set of practices and methodologies focused on manufacturing equipment performance improvement, TPM has matured into a comprehensive equipment-centric effort to optimize manufacturing productivity. TPM brings maintenance into focus as a necessary and vitally important part of the business. It is no longer regarded as a non-profit activity. TPM describes a synergistic relationship among all organizational functions, but particularly between production and maintenance, for continuous improvement of product quality, operational efficiency, productivity, and safety. TPM is an indispensable strategic initiative to meet customer's demands on price, quality, and lead-times. Willmott (1994) portrays TPM as a relatively new and practical application of TQM and suggests that TPM aims to promote a culture in which operators develop 'ownership' of their machines, learn much more about them, and in the process realize skilled trades to concentrate on problem diagnostic and equipment improvement projects.

From a lean manufacturing perspective, improved efficiency and profitability can be sought by increasing value within an organization through the elimination of waste. TPM focuses on systematic identification and elimination of waste, inefficient operation cycle time, and quality defects in manufacturing and processes (McCarthy, 2004). TPM is based on teamwork and provides a method for the achievement of world class levels of overall equipment effectiveness (OEE) through people and not through technology or systems alone. TPM is an approach to equipment management that involves employees from both production and maintenance departments through cross-functional teams. TPM is not a maintenance specific policy; it is a culture, a philosophy, and a new attitude towards maintenance. An effective TPM program can facilitate enhanced organizational capabilities across a variety of dimensions (Wang, 2006). Strategic TPM implementation success factors like top management leadership and involvement, traditional maintenance practices, and holistic TPM implementation initiatives can contribute towards effecting significant improvements in manufacturing performance (Ahuja and Khamba, 2008c).

The TPM literature offers a number of definitions for Total Productive Maintenance:

- TPM is an innovative approach to maintenance that optimizes equipment effectiveness, eliminates breakdowns, and promotes autonomous maintenance by operators through day-to-day activities involving the total workforce (Nakajima, 1989);
- TPM is a partnership between maintenance and production function organizations to improve product quality, reduce waste, reduce manufacturing cost, increase equipment availability, and improve organization's state of maintenance (Rhyne, 1990);
- TPM is a maintenance improvement strategy that involves all employees in the organization and includes everyone from top management to the line employee and encompasses all departments including maintenance, operations, design engineering, project engineering, inventory and stores, purchasing, accounting finances, and plant management (Wireman, 1990);
- TPM is a production-driven improvement methodology that is designed to optimize equipment reliability and ensure efficient management of plant assets (Robinson and Ginder, 1995);
- TPM is a program that addresses equipment maintenance through a comprehensive productive-maintenance delivery system covering the entire life cycle of equipment and involving all employees from production, maintenance personnel to top management (McKone *et al.* 1999); and
- TPM is about communication; it mandates that operators, maintenance people and engineers collectively collaborate and understand each other's language (Witt, 2006).

In 1971, Japan Institute of Plant Maintenance (JIPM) defined TPM (Nakajima, 1988; Heston, 2006), focusing mainly upon the production sector, as:

- TPM aims to maximize equipment efficiency (overall efficiency improvement);
- TPM aims to establish total system of PM, designed for the entire life of equipment;
- TPM operates in all sectors involved with equipment, including the planning, using and maintenance sector;
- TPM is based on participation of all members, from top management to frontline staff members; and
- TPM carries out PM through motivation management, *i.e.*, small-group activities.

However, as TPM outgrew the production department, to be implemented organization-wide, TPM definition has been subsequently modified as (Shirose, 1996):

- TPM aims to create a corporate system that maximizes the efficiency of production system (Overall Efficiency Improvement);
- TPM establishes a mechanism for preventing the occurrence of all losses on the front line and is focused on the end product, this includes systems for realizing 'zero accidents, zero defects and zero failures' in the entire life cycle of the production system;

- TPM is applied in all sectors, including the production, development and administration departments;
- TPM is based on the participation of all members, ranging from top management to frontline employees; and
- TPM achieves zero losses through overlapping small-group activities.

17.2 Evolution Towards TPM

The maintenance function has undergone a significant change in the last three decades. Equipment management has passed through many phases. The progress of maintenance concepts over the years is explained below.

1. *Breakdown maintenance (BM)*: this is the maintenance strategy, whereby repair/restoration is initiated after the equipment failure/stoppage or upon occurrence of severe performance decline. This maintenance strategy was primarily adopted in manufacturing organizations, worldwide, prior to the 1950s. In this strategy, machines are serviced only when repair is drastically required. This concept has the disadvantage of long unplanned stoppages, excessive damage, spare parts problems, high repair costs, excessive waiting and maintenance time, and high troubleshooting problems.
2. *Preventive maintenance (PM)*: this concept, introduced in 1951, is a kind of physical check-up of the equipment to prevent equipment breakdown and prolong equipment service life. PM comprises of maintenance activities that are undertaken after a specified period of time or amount of machine use. During this phase, the maintenance function is established and time based maintenance (TBM) activities are generally accepted. This type of maintenance relies on the estimated probability that equipment will break down or experience deterioration in performance in a specified time interval. The preventive work undertaken may include equipment lubrication, cleaning, parts replacement, tightening, and adjustment. The production equipment may also be inspected for signs of deterioration during preventive maintenance work.
3. *Predictive maintenance (Pd.M.)*: predictive maintenance is often referred to as condition based maintenance (CBM). In this strategy, maintenance is initiated in response to specific equipment condition or performance deterioration. The diagnostic techniques are deployed to measure physical condition of the equipment such as temperature, noise, vibration, lubrication, and corrosion. When one or more of these indicators reach a predetermined deterioration level, maintenance initiatives are undertaken to restore the equipment to desired condition. This means that equipment is taken out of service only when direct evidence exists that deterioration has taken place. Predictive maintenance is premised on the same principal as preventive maintenance although it employs a different criterion for determining the need for specific maintenance activities. The additional benefit comes from

the need to perform maintenance when imminent and not after the passage of a specified period of time.

4. *Corrective maintenance (CM)*: this is a strategy, introduced in 1957, in which the endeavor to prevent equipment failures is further expanded to be applied to improvement of equipment so that equipment failures can be eliminated (improving the reliability) and equipment can be easily maintained (improving equipment maintainability). The primary difference between corrective and preventive maintenance is that a problem must exist before corrective actions are taken. The purpose of corrective maintenance is to improve equipment reliability, maintainability, safety, and design weaknesses (material, shapes). The corrective maintenance strategies aim to reduce deteriorations, failures for ensuring maintenance-free equipment. Maintenance information, obtained from CM, is useful for maintenance prevention for the new generation equipment and improvement of existing manufacturing facilities.
5. *Maintenance prevention (MP)*: introduced in the 1960s, this is an activity wherein the piece of equipment are designed such that they are maintenance-free and an ultimate ideal condition of 'what the equipment and the line must be' is achieved. In the development of new equipment, MP initiatives must start at the design stage and strategically aim at ensuring reliable equipment, easy to care for and user friendly, so that operators can easily manage, adjust, and run it. Maintenance prevention often functions using experience from earlier equipment failures, product malfunctionings, feedback from production areas, customers, and marketing functions to ensure hassle free operation for existing and new production systems.
6. *Reliability centered maintenance (RCM)*: reliability centered maintenance was founded in 1960s and was primarily oriented towards maintaining airplanes and used by aircraft manufacturers, airlines, and government facilities. RCM is a structured, logical process for developing or optimizing the maintenance requirements of a physical resource in its operating context to realize its 'inherent reliability', where 'inherent reliability' is the level of reliability which can be achieved with an effective maintenance program. RCM is a process used to determine the maintenance requirements of physical asset in its operating context by identifying the functions of the asset, the causes of failures, and the effects of the failures. The various tools employed for affecting maintenance improvement include failure mode and effect analysis (FMEA), failure mode effect and criticality analysis (FMECA), physical hazard analysis (PHA), fault tree analysis (FTA), optimizing maintenance function (OMF), and hazard and operability (HAZOP) analysis.
7. *Productive maintenance (Pr.M)*: productive maintenance means the most economic maintenance that raises equipment productivity. The purpose of productive maintenance is to increase productivity of an enterprise by reducing total cost of equipment over the entire life cycle. The key characteristics of this maintenance philosophy are equipment reliability and maintainability focus. The maintenance strategy involving all activities to improve equipment productivity by performing preventive maintenance,

corrective maintenance, and maintenance prevention throughout the life cycle of equipment is called productive maintenance.

8. *Computerized maintenance management systems (CMMS)*: computerized maintenance management systems assist in managing a wide range of information on maintenance workforce, spare-parts inventories, repair schedules, and equipment histories. It may be used to plan and schedule work orders, expedite dispatch of breakdown calls, and manage the overall maintenance workload. CMMS can be deployed to automate the PM function and to assist in the control of maintenance inventories and the purchase of materials. CMMS has the potential to strengthen reporting and analysis capabilities. The capability of CMMS to manage maintenance information contributes to improved communication and decision-making capabilities within the maintenance function. Accessibility of information and communication links on CMMS ensures improved maintenance responsiveness, better communication of repair needs and work priorities, and improved coordination through closer working relationships between maintenance and production.
9. *Total productive maintenance (TPM)*: total productive maintenance is a unique Japanese philosophy, which has been developed based on productive Maintenance concepts and methodologies. This concept was first introduced by M/s Nippon Denso Co. Ltd. of Japan, a supplier of M/s Toyota Motor Company, Japan in 1971. TPM is an innovative approach to maintenance that optimizes equipment effectiveness, eliminates breakdowns, and promotes autonomous maintenance by operators through day-to-day activities involving total workforce.

TPM initiative is targeted to enhance competitiveness of the enterprises and encompasses a powerful structured approach to change the mind-set of employees, thereby making a visible change in work culture of the organizations. TPM seeks to engage all levels and functions in the organizations to maximize overall effectiveness of production facilities. TPM is a world class manufacturing (WCM) initiative that seeks to optimize the effectiveness of manufacturing equipment. Whereas maintenance departments are the traditional center of preventive maintenance programs, TPM seeks to involve workers from all departments and levels, including plant-floor operators to senior executives, to ensure effective equipment operation.

17.3 Need of TPM

The rapidly changing needs of modern manufacturing and ever increasing global competition has emphasized the need for re-examination of the role of improved maintenance management towards enhancing an organization's competitiveness. This has provided the impetus to leading organizations world-wide to adopt effective and efficient maintenance strategies such as CBM, RCM, and TPM over the traditional firefighting reactive maintenance approaches. A strategic approach

to improve performance of maintenance activities is to adapt and implement effectively strategic TPM initiatives in the manufacturing organizations.

TPM harnesses participation of all the employees to improve production equipment availability, performance, quality, reliability, and safety. TPM endeavors to tap the 'hidden capacity' of unreliable and ineffective equipment. TPM capitalizes on proactive and progressive maintenance methodologies and calls upon knowledge and co-operation of operators, equipment vendors, engineering, and support personnel to optimize machine performance, thereby resulting in elimination of breakdowns, reduction of unscheduled and scheduled downtime, improved utilization, higher throughput, and better product quality. The bottom-line achievements of successful TPM implementation initiatives in an organization include lower operating costs, longer equipment life and lower overall maintenance costs.

The following aspects necessitate need for implementing TPM in the contemporary manufacturing scenario:

- To become world class, satisfy global customers and achieve sustained organizational growth;
- Need to change and remain competitive;
- Need to monitor critically and regulate work-in-process (WIP) out of 'Lean' production processes owing to synchronization of manufacturing processes;
- Achieving enhanced manufacturing flexibility objectives;
- To improve organization's work culture and mindset;
- To improve productivity and quality;
- Tapping significant cost reduction opportunity regarding maintenance related expenses;
- Minimizing investments in new technologies and maximizing return on investment ROI;
- Ensuring appropriate manufacturing quality and production quantities in JIT manufacturing environment;
- Realizing paramount reliability and flexibility requirements of the organizations;
- Regulating inventory levels and production lead-times for realizing optimal equipment available time or up-time;
- Optimizing life cycle costs for realizing competitiveness in the global market-place;
- To obviate problems faced by organizations in form of external factors like tough competition, globalization, increase in raw material costs and energy cost;
- Obviating problems faced by organizations in form of internal factors like low productivity, high customer complaints, high defect rates, non-adherence to delivery time, increase in wages and salaries, lack of knowledge, skill of workers, and high production system losses;
- Ensuring more effective use of human resources, supporting personal growth and garnering of human resource competencies through adequate training and multi-skilling;

- To liquidate the unsolved tasks (breakdown, setup time and defects);
- To make the job simpler and safer; and
- To work smarter and not harder (improve employee skill).

17.4 Basic Elements of TPM

TPM is an important world-class manufacturing program introduced during the quality revolution. TPM seeks to maximize equipment effectiveness throughout the lifetime of equipment. It strives to maintain equipment in optimum condition in order to prevent unexpected breakdowns, speed losses and quality defects occurring from process activities. There are three ultimate goals of TPM: zero defects, zero accident, and zero breakdowns. Nakajima (1988) suggests that equipments should be operated at 100% capacity 100% of the time. The benefits arising from TPM can be classified in six categories including productivity (P), quality (Q), cost (C), delivery (D), safety (S) and morale (M). TPM has been envisioned as a comprehensive manufacturing strategy to improve equipment productivity. Benchmarking on OEE, P, Q, C, D, S and M can enable an organization to realize zero breakdown, defect, machine stoppage, accidents, and pollution, which serve as an ultimate objective of TPM. The strategic elements of TPM include cross-functional teams to eliminate barriers to machine uptime, rigorous preventive maintenance programs, improved maintenance operations management efficiency, equipment maintenance training to the lowest level, and information systems to support the development of imported equipment with lower cost and higher reliability. Similar to TQM, TPM is focused on improving all the big picture indicators of manufacturing success. TPM implementation requires a long-term commitment to achieve the benefit of improved OEE through training, management support and teamwork.

Figure 17.1 shows the framework of TPM implementation and depicts tools used in TPM implementation program with potential benefits accrued and targets sought. TPM initiatives as suggested by Japan Institute of Plant Maintenance (JIPM) involve an eight pillar implementation plan that results in substantial increase in labor productivity through controlled maintenance, reduction in maintenance costs, and reduced setup and downtimes. The basic principals of TPM are often called the pillars or elements of TPM. The entire edifice of TPM is built and stands on eight pillars. TPM paves the way for excellent planning, organizing, monitoring, and controlling practices through its unique eight pillar methodology involving: autonomous maintenance; focused improvement; planned maintenance; quality maintenance; education and training; safety, health and environment; office TPM; and development management (Rodrigues and Hatakeyama, 2006). The eight pillar Nakajima model of TPM implementation has been depicted in Figure 17.2, while Figure 17.3 shows maintenance and organizational improvement initiatives associated with the respective TPM pillars (Ahuja and Khamba, 2007).

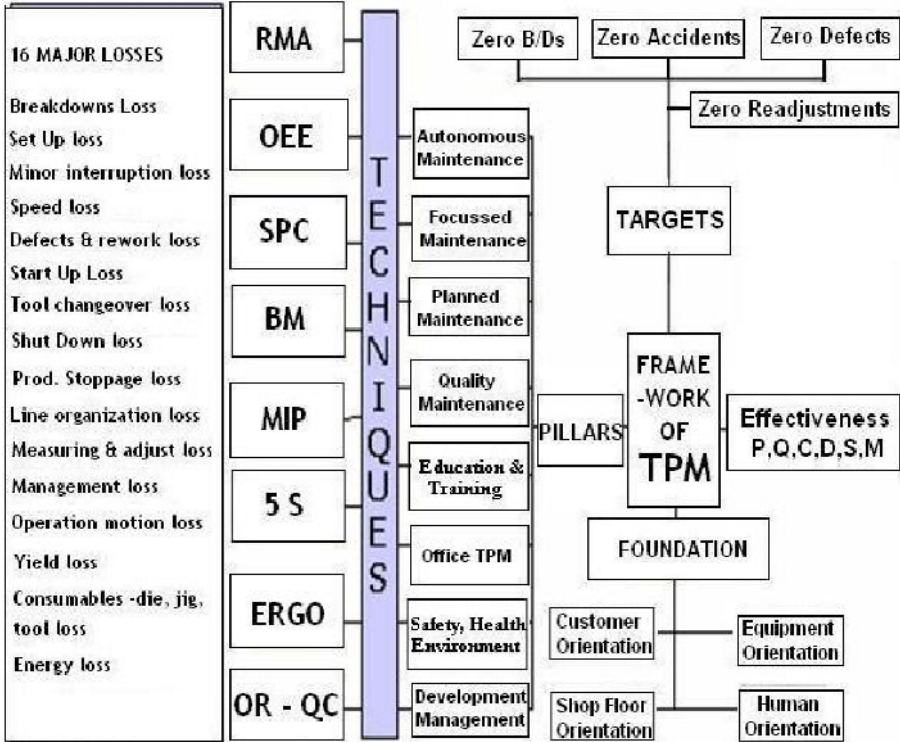


Figure 17.1. Framework of total productive maintenance

TPM initiatives aim at achieving enhanced safety, asset utilization, production capacity without additional investments in new equipment, human resources and continuing to lower the cost of equipment maintenance and improving machine uptime. It provides an effective way of deploying activities through its TPM promotion organization involving 100% of employees on a continuous basis. The main goal of an effective TPM program is to bring critical maintenance skilled trades and production workers together. Total employee involvement, autonomous maintenance by operators, small group activities to improve equipment reliability, maintainability, productivity, and continuous improvement (Kaizen) are the principles embraced by TPM. There are a variety of tools that are traditionally used for quality improvement. TPM uses the following tools among others to analyze and solve the equipment and process related problems: pareto analysis; statistical process control (SPC - control charts); problem solving techniques (brainstorming, cause-effect diagrams, and 5-M approach); team based problem solving; poka-yoke systems (mistake proofing); autonomous maintenance; continuous improvement; 5S; setup time reduction (SMED); waste minimization; benchmarking; bottleneck analysis; reliability, maintainability and availability (RMA) analysis; recognition and reward programs; and system simulation. TPM provides a comprehensive, life cycle approach to equipment management that minimizes equipment failures, production defects, and accidents. The objective is to improve continuously

production system availability and prevent degradation of equipment to realize maximum effectiveness. These objectives require strong management support as well as continuous use of work teams and small group activities to achieve incremental improvements.

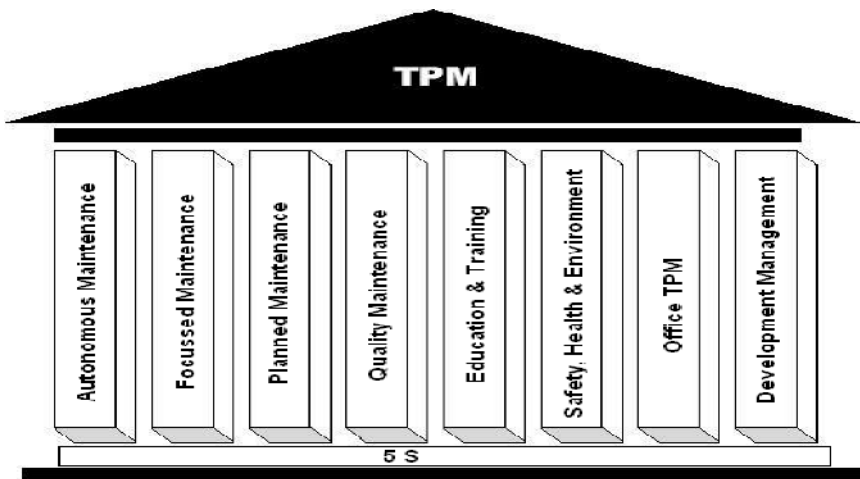


Figure 17.2. Eight pillar approach for TPM implementation (suggested by JIPM)

TPM employs OEE as the core quantitative metric for measuring the performance of a productive system. OEE has been widely accepted as an essential quantitative tool for measurement of productivity of manufacturing operations. The role of OEE goes far beyond the task of just monitoring and controlling. OEE measure is central to the formulation and execution of a TPM improvement strategy. It provides a systematic method for establishing production targets and incorporates practical management tools and techniques in order to achieve a balanced view of process availability, performance rate, and quality. OEE has been used as an impartial daily snapshot of the equipment and promotes openness in information sharing and a no-blame approach in handling equipment related issues. OEE is the measure of contribution of current equipment to the added value generation time, based on overall consideration of time, speed performance, and non-defective ratio of the equipment. The improvement of OEE is essential to drive a lean production system and is calculated by multiplying availability of equipment, performance efficiency of process and rate of quality products (Gregory, 2006).

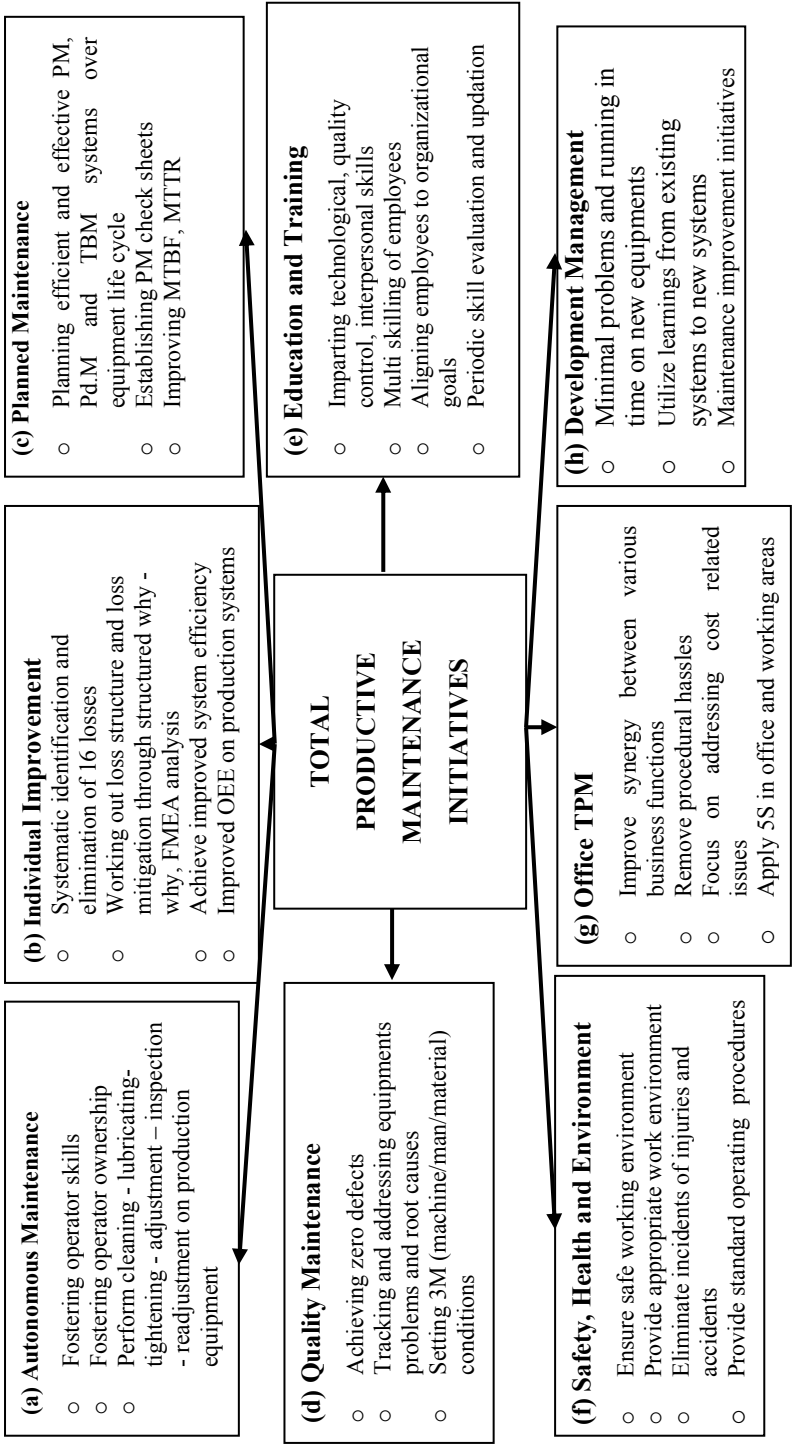


Figure 17.3. TPM initiatives associated with various pillars

OEE = Availability (A) × Performance Efficiency (P) × Rate of Quality (Q) (17.1)

where

Availability (A) = $\frac{\text{Loading Time} - \text{Downtime}}{\text{Loading Time}} \times 100$ (17.1a)

Performance Efficiency (P) = $\frac{\text{Processed Amount}}{\text{Operating Time/Theoretical Cycle Time}} \times 100$ (17.1b)

Rate of Quality = $\frac{\text{Processed Amount} - \text{Defect Amount}}{\text{Processed Amount}} \times 100$ (17.1c)

TPM has the standards of 90% availability, 95% performance efficiency and 99% rate of quality parts. An overall 85% of OEE is considered as world class and a benchmark for others. TPM seeks to improve the OEE, which is an important performance indicator, used to measure success of TPM in an organization. In the initial stages, TPM initiatives focus upon addressing six major losses, which are considered significant in affecting the efficiency of the production system. The six major losses include: equipment breakdown losses; setup and adjustment losses; idling, minor stoppage losses; reduced speed losses; defect and rework losses; and startup losses. TPM endeavors to increase efficiency by rooting out losses that sap productive efficiency. The calculation of OEE by considering six major production losses has been depicted in Figure 17.4 (McKellen, 2005). Using OEE metrics and establishing a disciplined reporting system help an organization to focus on parameters critical to its success.

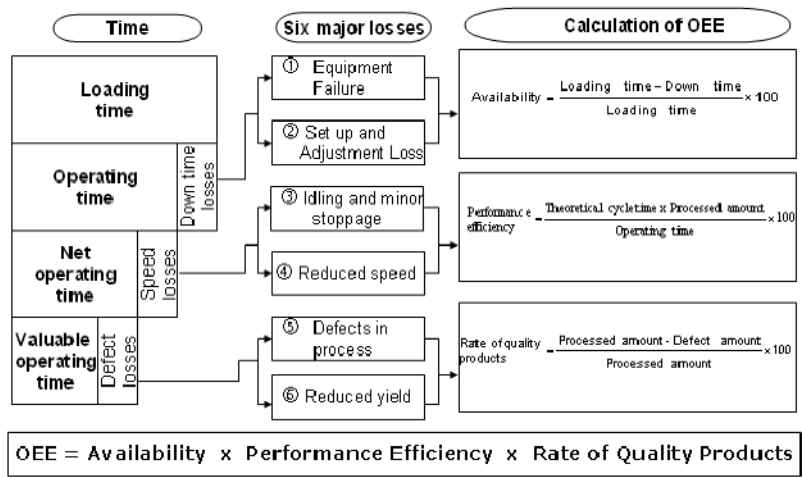


Figure 17.4. Calculation of overall equipment effectiveness based on six major losses

In the quest to achieve world class manufacturing, organizations have been relying upon exhaustive analysis of manufacturing systems in order to ascertain inefficiencies, weaknesses hampering the production system performance. It has been observed that other than equipment related losses, losses affecting human performance, energy, and yield inefficiencies also need to be investigated and addressed appropriately for achieving world class performance. For this purpose, 16 major losses have been identified to be severely impeding the manufacturing performance. These losses have been categorized into four categories, which include seven major losses impeding equipment efficiency (failure losses, setup/adjustment losses, reduced speed losses, idling/minor stoppage losses, defect/rework losses, startup losses, and tool changeover losses); losses impeding machine loading time (planned shutdown losses); five major losses impeding human performance (distribution/logistic losses, line organization losses, measurement/adjustment losses, management losses, and motion related losses) and three major losses impeding effective use of production resources (yield losses, consumable – jig/tool/die losses, and energy losses) (Shirose, 1996). The calculation of OEE by considering impact of sixteen major losses on production system has been depicted in Figure 17.5.

OEE metric offers a starting point for developing quantitative variables for relating maintenance measurement to corporate strategy. OEE measure provides a strong impetus for introducing a pilot and subsequently an organization-wide TPM program. OEE can be used as an indicator of reliability of a production system. OEE is a productivity improvement process that starts with management awareness of total productive manufacturing and their commitment to focus the factory workforce on training in teamwork and cross-functional equipment problem solving. Forming cross-functional teams to solve the root causes/problems drive the greatest improvements and generate real bottom-line earnings. A comparison between expected and current OEE measures provides much needed impetus for manufacturing organizations to improve maintenance policy and effect continuous improvements in the manufacturing systems.

17.5 Roadmap for TPM Implementation

Lycke and Akersten (2000) have suggested that TPM is a highly structured approach and careful, thorough planning and preparation are keys to successful organization-wide implementation of TPM and so is senior management's understanding and belief in the concept. One of the most significant elements of the TPM implementation process is that it is a consistent methodology for continuous improvement. TPM is a long-term process, not a quick fix strategy for today's manufacturing problems. The organizations across the world have been struggling for a long time to evolve the best possible set of strategies for successful implementation of TPM. However, TPM experts and practitioners around the world have now acknowledged problems regarding a cookbook-style TPM in the organizations due to factors like: highly variable skills associated with workforce under different situations; age differences of workgroups; varied complexities of production systems and equipments; altogether different organization cultures,

objectives, policies and environments; and differences in the prevailing status of maintenance competencies.

Although “there is no single right method for implementation of a TPM program” and there has been “a complexity and divergence of TPM programs adopted throughout industry”, it is clear that a structured implementation process is an identified success factor and a key element of TPM programs (Bamber *et al.* 1999). In order to introduce successfully principles and practices of TPM, an elaborative and structured TPM implementation methodology is necessary to facilitate organizations to effect a smooth transition from current state to the desired world class manufacturing performance levels. There have been many approaches suggested by different practitioners and researchers for implementing TPM in different organizations, having varying work environments and organizational objectives for garnering strategic manufacturing competencies.

Nakajima has also outlined a 12 step TPM methodology involving 4 phases of TPM implementation (Nakajima, 1988; Shirose, 1996). These 12 steps support basic developmental activities, which constitute minimal requirements for the development of TPM. The various steps involved in the TPM implementation methodology have been depicted in Table 17.1.

Naguib (1993) has proposed a five phase roadmap for TPM implementation which includes: an awareness program to obtain management commitment and support; restructuring of manufacturing organization to integrate maintenance in production modules; planning maps to cover TPM activities related to equipment effectiveness, maintenance management system, and workplace environment enhancements; workforce competencies improvements; an implementation process based on cross-functional, multi-skilled, self-directed teams; and an assessment process to ‘close loop’ the implementation process and define directions for continuous improvements.

Another simplified Western approach involving ‘Five Pillar Model’ proposed by Steinbacher and Steinbacher (1993) has been presented in Figure 17.6. TPM implementation process, at the highest level, requires initialization, implementation, and institutionalization. In this model, ‘Training and Education’ is an integral element of all other pillars rather than a stand-alone pillar as depicted in the Nakajima model.

Pirsig (1996) has emphasized seven unique broad elements and four main themes in any TPM implementation program. The key themes in TPM implementation program include: training, decentralization, maintenance prevention, and multi-skilling, while the broad elements include: asset strategy, empowerment, resource planning and scheduling, systems and procedures, measurement, continuous improvement teams and processes. The inter-relationship between TPM themes and broad elements has been depicted in Figure 17.7.

A	AVAILABILITY	Last 3 Years Trend	Period I	Period II	Period III
1	Equipment Failure Loss				
1A	Parts Failure				
1B	Parts Adjustment				
2	Setup and Adjustment loss				
3	Tool Change Loss				
4	Startup Loss				
5	Minor Stoppage Loss				
6	Reduced Speed Loss				
7	Defect and Rework Loss				
8	Scheduled Downtime				
8A	Cleaning and Checking				
8B	Planned Maintenance				
8C	Meetings				
9	Management Loss				
9A	Waiting For Spares and Tools				
9B	Waiting For Instruction				
9C	Waiting For Material				
9D	Waiting For Men				
9E	Waiting For Power				
9F	Waiting For Inspection				
Total Non-Utilized Hours (a)					
Hours Available (b)					
Availability = $(b - a) / b$					
B	PERFORMANCE EFFICIENCY				
Output Achieved for $(b - a) = c$					
Standard Output for $(b - a) = d$					
Performance Efficiency = c / d					
C	QUALITY				
Quantity Produced (e)					
Quantity under RT (f)					
Quantity Scrapped (g)					
Rate of Good Components Produced = $\{e - (f - g)\} / e$					
OEE = A x B x C					

Figure 17.5. Measurement of overall equipment effectiveness

Table 17.1. Twelve step TPM implementation methodology

Phase of implementation	TPM implementation steps	Activities involved
Stage preparation	1. Declaration by top management decision to introduce TPM	• Declare TPM introduction at in-house seminar • Carried in organization magazine
	2. Launch education and campaign to introduce TPM	• Managers: trained in seminar/camp at each level • General employees: seminar meetings using slides
	3. Create organizations to promote TPM	• Create organizational heierarchy for TPM program • Constitute committees and sub-committees
	4. Establish basic TPM policies and goals	• Benchmarks and targets evolved • Prediction of effects
	5. Formulate master plan for TPM development	• Develop step-by-step TPM implementation plan • Framework of strategies to be adopted over time
Preliminary implementation	6. Hold TPM kick-off	• Invite suppliers, related companies, affiliated companies

Table 17.1. (continued)

TPM Implementation	7. Establishment of a system for improving the efficiency of production system	● Pursuit of improvement of efficiency in production department
	8. Improve effectiveness of each piece of equipment	● Project team activities and small group activities (SGA) at production centers
	9. Develop an autonomous maintenance (AM) program	● Step system, diagnosis, qualification certification
	10. Develop a scheduled maintenance program for the maintenance department	● Improvement maintenance, periodic maintenance, predictive maintenance
	11. Conduct training to improve operation and maintenance skills	● Group education of leaders and training members
	12. Develop initial equipment management program level	● Development of easy to manufacture products and easy to operate production equipment
	13. Establish quality maintenance organization	● Setting conditions without defectives, and its maintenance and control
	14. Establish systems to improve efficiency of administration and other indirect departments	● Support for production, improving efficiency of related sectors
	15. Establish systems to control safety, health and environment	● Creation of systems for zero accidents and zero pollution cases
Stabilization	16. Perfect TPM implementation and raise TPM performance	● Sustaining maintenance improvement efforts ● Challenging higher targets ● Applying for PM awards

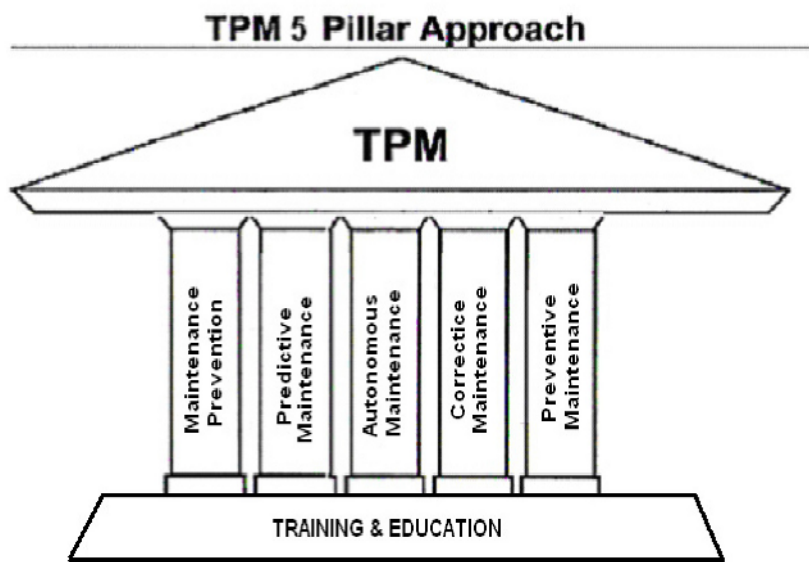


Figure 17.6. Steinbacher and Steinbacher model of TPM implementation

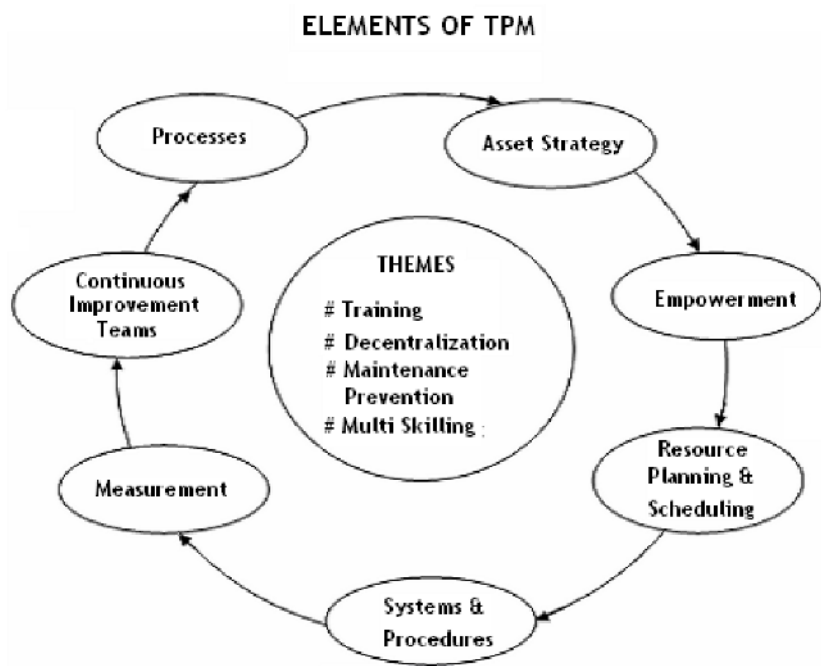


Figure 17.7. Pirsig model of TPM implementation

17.6 An Ideal TPM Methodology

An Ideal TPM Methodology (ITPMM) for manufacturing organizations has been categorized into three phases namely: introduction phase, TPM initiatives implementation phase and standardization phase. The initiatives associated with respective phases of ITPMM have been described in Figure 17.8.

The sequence of TPM implementation events can be modified depending on the needs of different organizations. ITPMM provides more capability of customization. It can be modified to meet the needs of the enterprises attempting to implement TPM. ITPMM supports the user to implement TPM in any time frame considered beneficial to the enterprise.

17.6.1 Introduction Phase (Phase I)

This is the first step for ensuring smooth implementation of TPM in manufacturing organization. It has been observed that most of the failures regarding TPM implementation programs arise on account of poor planning and false start-ups. This phase involves careful planning and deployment of prerequisites for successfully managing TPM initiatives in the organization and is crucial to success of TPM implementation.

The introduction phase initiatives help to address concerns of employees towards TPM implementation, align employees towards goals of organization, and ensure development of an effective roadmap for TPM implementation, thereby creating a favorable environment in the organization. The success of an organization regarding fruitful TPM implementation is primarily dependent upon its ability to implement activities mentioned in Phase I. The detailed description of initiatives to be deployed effectively by the manufacturing organization has been presented here.

17.6.1.1 Top Management Commitment

The successful deployment of a strategic TPM implementation plan requires top management support, commitment, involvement, and requires an aggressive and supportive management team. Top management can significantly contribute by more than just allowing TPM to be implemented at the organization, but can actually be a part of the driving force behind TPM implementation. Management needs to have a strong commitment to the TPM implementation program and should go all-out for evolving mechanisms for multi-level communication to all employees, explaining importance and benefits of the program, whole heartedly propagating TPM benefits to the organization and employees by linking TPM to overall organizational strategy and objectives. Management must make sincere efforts to ensure union buy-in for successful management of a TPM program in the organization. The unions need not be treated as an adversary in affecting workplace transformations.

The management contributions towards successful TPM implementations can include: revising business plans to include TPM goals, affecting appropriate cultural transformations in organizational culture, building strong success stories for promoting motivation for TPM, communicating TPM goals, providing

adequate financial resources for affecting business improvements, promoting and nurturing cross-functional team culture, providing training and skill enhancements for production and maintenance workers, evolving appropriate reward and incentive mechanisms for promoting continuous improvement, ensuring total employee involvement, supporting changes and improvements at the workplace, removing barriers related to middle level management, and enhancing inter-department synergy.

PHASE III Standardization Phase	Sustain TPM initiatives							
	Deploy lean manufacturing practices							
	Deploying key performance indicators for assessing manufacturing performance							
PHASE II TPM initiatives implementation phase								
Autonomous maintenance	Focused improvement	Planned maintenance	Quality maintenance	Office TPM	Safety, health and environment	Development management	Tool management	Maintenance benchmarking
	Visual workplace							
	Computerized maintenance management system (CMMS)							
	Inculcate teamworking culture							
	Training and multi-skilling for TPM							
	Continuous improvement and Kaizen							
	Employee empowerment							
	Managing successful organizational cultural transformation							
PHASE I Introduction Phase	Top management commitment							

Figure 17.8. Ideal TPM methodology (ITPMM) for manufacturing organizations

The first course of action is to establish strategic directions for TPM. This can be achieved by evolving appropriate TPM policy and a master plan towards TPM implementation. Developing TPM goals and a master plan is the process of evolving a desired future condition and a scheme, which helps to achieve it. The master plan developed in the activity that guides the process of TPM implementation. At this juncture, the structured TPM secretariat should be evolved and TPM promotion, steering committees formed and champions selected with their full commitment in this process to ensure the success of TPM implementation. Finally, management must ensure that laid out procedures are holistically implemented in the organization. Management must periodically review progress of the TPM implementation program against the laid out master plan, mission and make suitable amendments for ensuring effective implementation of TPM program.

17.6.1.2 Managing Successful Organizational Cultural Transformation

The biggest challenge before management is to be able to make radical transformation in an organization's culture for ensuring overall employee participation towards manufacturing performance improvement through TPM initiatives. Creating cultural transformation is the process of changing existing culture to adapt effectively to harness employee competencies to implement TPM. Also, it transforms the culture into competent policies and reward mechanisms to facilitate TPM implementation. The management should endeavor to develop favorable policies and reward systems, building sense of ownership for operators, improving communication and trust, and conducting training and education for TPM implementation. The success of an organization is dependent upon the ability of management to overcome strategically barriers affecting TPM implementation. Moreover, many other strategic initiatives for motivating and aligning employees to organizational goals can also be successfully deployed in the organizations. These include evolving mechanisms for employee empowerment, recognition of efforts made by employees towards organizational performance enhancement, making sincere efforts for improving skill and knowledge base of all employees, and promoting cross-functionality between various organization functions. The strategic issues for achieving cultural transformations in the organization should also include effectively planning changeover, evolving strategic action plans, and allocation resources for implementing changes at workplace for ensuring total employee involvement for successful TPM implementation.

17.6.1.3 Employee Empowerment

Employee empowerment (employee involvement or workplace democracy) means 'the extent to which employees producing a product or offering a service have a sense of controlling their work, receiving information about their performance, and being rewarded for affecting performance enhancement in the workplace. Total employee involvement and integration is a pre-requisite to successful TPM implementation and can be ensured by enhancing employee competencies towards the jobs, evolving a culture of equipment and system ownership by the employees, adequate employee counseling, union buy-in, effective appropriate suggestions schemes, and deploying encouraging and safe work environment in the

organization. One of the essential principals of TPM is encouraging operators to assume more responsibility and authority for decisions affecting their production equipment.

Employee involvement through strategic TPM programs can lead to: greater acceptance of decisions; commitment to improvement ideas; understanding of objectives; fulfillment of psychological needs and intrinsic satisfaction; team identity; cooperation and coordination; effectiveness of group decisions and better conflict resolution; employee participation in work decisions; consultative participation; informal participation; employee ownership; representative participation. The manufacturing enterprise must ensure enhanced employee involvement by promoting quality circles in the organization and offer job enrichment for improved employee performance and job satisfaction. Self directed work (SDW) teams lead to improved employee attitudes and behavior thereby resulting in reduced turnover and absenteeism. The employee ownership has also been revealed to provide positive impact on organizational productivity and employee attitudes. After introduction of autonomous maintenance initiatives, operators take care of machines by themselves without being ordered to. With the realization of zero breakdowns, zero accidents, and zero defects, operators get new confidence in their abilities and organizations also understand the importance of employee contributions towards realization of manufacturing performance enhancements.

TPM program emphasizes unique basic techniques and strategic, human resource oriented practices. The manufacturing organization can contribute in this regard by ensuring multi-skilling of all employees for ensuring employee involvement (EI). The types of skills frequently identified as being necessary for effective employee involvement are group decision making, problem solving skills, leadership skills, skills involving understanding of business (finance, accounting, quality control), statistical analysis skills, team building skills, and job related skills. The organization can institutionalize efficient compensation, rewards and appraisal systems for ensuring total employee involvement by adopting individual incentive plans, team incentives, profit sharing, gain sharing, employee stock ownership plan, knowledge – skill based incentives, as well as, non-monetary rewards, incentives like recognitions, felicitation, and honors, *etc.* Moreover, information sharing, power sharing, personnel policies and practices (employment security, hiring mechanisms, flexible-timing, suggestion schemes) can also lead to improving employee involvement. TPM implementation also helps to foster motivation in the workforce through adequate empowerment, training, and felicitations, thereby enhancing the employee participation towards realization of organizational goals and objectives (Ahuja and Khamba, 2008a). The other benefits include favorable changes in the attitude of the operators, achieving goals by working in teams, sharing knowledge and experience, and workers getting a feeling of owning the production facilities. Although management needs to assume a leadership role in TPM implementation, they must also allow equipment operators to take a prominent role in the development and implementation of TPM.

17.6.1.4 Continuous Improvement and Kaizens

Continuous improvement (CI) is associated with continuous quality and productivity improvement and is referred to in the broader context of waste elimination, Just-in-time manufacturing, and TPM. The success of TPM program is extensively dependent upon competencies and motivation of the workforce to affect significant improvement in production systems through CI and Kaizens. It has been recognized that higher quality levels can only be attained through continuous improvements to both products and processes. Effective TPM deployment requires senior management's commitment to Kaizen for sustained changes in organizational culture. The management's job is mainly comprised of two elements: 'maintenance', *i.e.*, maintaining current performance standards, and 'improvement', *i.e.*, raising performance standards. The TPM program calls for involving everybody in the organization to contribute effectively towards growth and development of the organization. Kaizen means ongoing improvement involving everyone – top management, managers and workers.

CI refers to continuous improvement of processes and systems, which in term manifests as CI on many fronts, such as CI of productivity, quality, cost, schedule, production, process flow, and so forth. CI can be considered as a system to ensure that incremental improvements do not 'just happen', but they can be seriously pursued. CI and Kaizens assume significance, since the implementation of new ideas or changes, big or small, have huge potential to contribute to organizational objectives. Continuous improvement plans and Kaizens should be holistically adopted for affecting significant organizational improvements including inventory reductions, reducing setup or changeover times, ensuring improved housekeeping and cleanliness, improving safety and hygiene at workplace, deploying Poka-Yoke initiatives at workplace, addressing equipment related improvements, conducting autonomous checks of abnormalities at the workplace, and deploying visual controls at the workplace.

The organization must evolve effective employee suggestion schemes to motivate employees to come forward and contribute towards the success of the enterprise. The organization can effectively contribute towards building a culture of continuous improvement and Kaizen at workplace by demonstrating their willingness to accept changes/improvements at the workplace and encouraging the suggestions given by employees. The organization must contribute in this regards by providing appropriate training for improving skill, knowledge towards CI, since CI is closely interlinked with skill and the knowledge base of employees.

17.6.1.5 Training and Multi-skilling for TPM

The core values and competencies of employees provide the foundation for TPM implementation. Employee's skills must be nurtured to meet the needs of their expanded roles. Therefore, adequate training and education of employees at all levels should be treated as a strategic initiative for successful TPM implementation. To keep up with changes in technology and equipment, craftworker training needs to be an ongoing process. The employees must not only be provided with technical job related skills and competencies (operating, maintaining and repairing new equipment technologies, preventive maintenance techniques, test equipment operation, calibration, fault analysis, safety training,

etc.), but also need to be well equipped with quality improvement and behavioral training for changing the mind set of employees from 'I operate, you inspect, you maintain' to 'I produce, I inspect, I maintain'. The training objectives must include systematic development of knowledge, skills, and attitude required by an individual to perform adequately the job responsibilities. The strategic skill enhancement training methodologies suggested for manufacturing organization has been outlined in Figure 17.9.

Top management's responsibility in this regard becomes identification of training needs, setting training targets, evolving appropriate training plans, preparation of training schedules, designing of training programs and material, providing JIT training, and evaluation of training effectiveness. The first-line maintenance crew and supervisors should be provided with state-of-the art technical skills and competencies such as technical knowledge of crafts, preventive maintenance methods, maintenance scheduling methods, and tools for planning and estimating maintenance work requirements. Maintaining a high level of skill on the part of maintenance craftworkers and supervisors helps to improve the quality of maintenance. Improved knowledge of planning and scheduling methods also supports the efficient use of maintenance resources. The top management must endeavor to train and develop employee competencies by updating their skill, knowledge, and attitude to enable higher productivity and achieve highest standards of quality, to eliminate product defects, equipment failures (breakdowns) and accidents, to develop multi-skilled workforce, and to create a sense of pride and belonging among all employees.

17.6.1.6 Inculcate Teamworking Culture

An effective TPM program calls for deployment of teams for improving equipment performance through and critical investigation of current and potential equipment problems. An important structure for employee involvement in TPM is cross functional teams (CFT). Teams help to break down the barriers that are inherent in the traditional approach to maintenance. Teams also help to identify problems and suggest new approaches for elimination of problems, introduce new skills that are needed, initiate training programs, and define TPM processes. Cross functional teams may involve participation from maintenance, R & D, process planning, production, and engineering that work together on an ongoing basis or temporary groups formed to address specific problems. The technical skills of engineers and experience of maintenance workers and equipment operators are communicated through these teams. One key strategy in effective implementation of workgroups is ensuring management's support to the efforts to drive CI in the team environment. Team leadership should include encouragement, facilitating and maintaining order, and help with decision-making. The organization must work progressively for promoting smooth functioning of cross functional teams, autonomous work teams (AWT), and problem solving groups (PSG).

Maintainability improvement and maintenance prevention are two key team-based TPM activities. Maintainability improvement teams work to improve the ways in which maintenance is performed. Maintenance, production workers, craft workers, and engineers work together to identify and correct poor equipment conditions like: difficult to locate, clean, inspect, handle, operate, and maintain

situations. This allows a wide range of improvements to be considered and deployed as appropriate. Maintainability improvements should result in increased maintenance efficiency and reduced maintenance time. Maintenance prevention teams work to improve equipment performance through enhanced equipment design. The maintenance function works with the engineering department during early stages of equipment design and allows the continuous improvement teams to design and install equipment that is easy to maintain and operate. Over the long term, efforts of maintenance improvement teams should result in improved equipment availability and reduced maintenance costs.

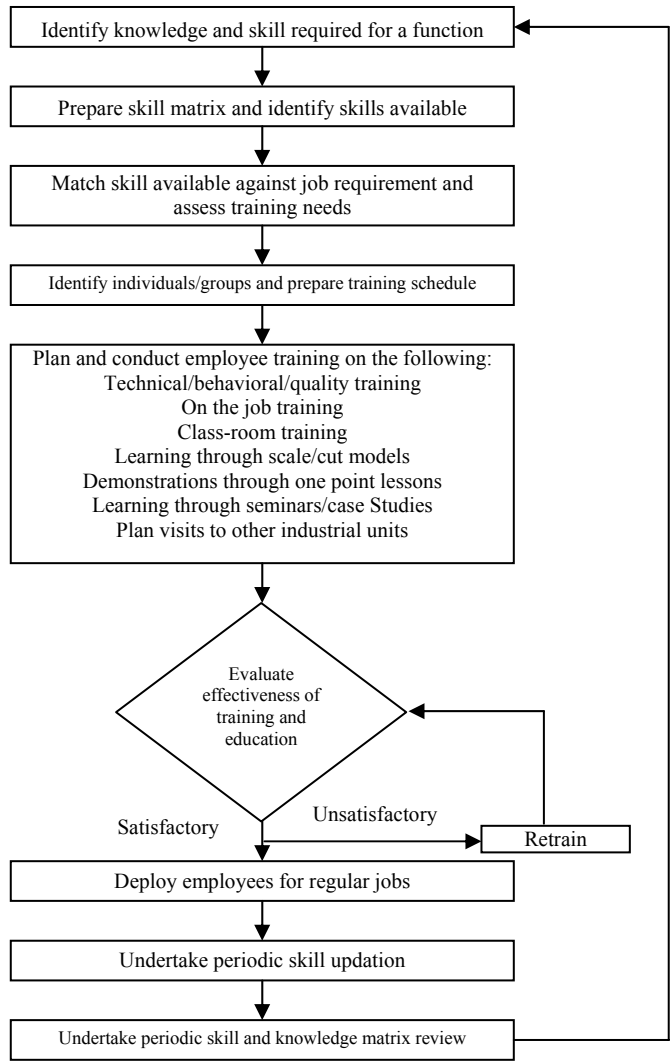


Figure 17.9. Skill enhancement training methodologies

17.6.1.7 Computerized Maintenance Management System (CMMS)

One of the major problems affecting effective maintenance improvement initiatives in manufacturing organizations is poor performance of industry in recording maintenance patterns and behaviors, poor spare part management, and ineffective deployment of maintenance improvements on future production systems. This can be attributed to inadequate utilization of CMMS and minimal role of information technology (IT) to sort out maintenance related problems. Thus, deployment of effective CMMS for effecting significant improvement in maintenance performance is strongly recommended in the manufacturing industry. CMMS have been the central focus for equipment management since the 1980s and they have been viewed as keys to achieve maintenance efficiency. CMMS must be holistically deployed to assist in managing a wide range of information on maintenance workforce, spare-parts inventories, repair schedules and equipment histories.

CMMS can be used to facilitate many functions including planning and scheduling work orders, expediting dispatch of breakdown calls, managing overall maintenance workload, tracking maintenance activities, costs, equipment failures, inventory control systems, and asset management capabilities. CMMS can be used to automate preventive maintenance function and control of maintenance inventories and purchase of materials. The LAN, WAN, and office computing technology allows CMMS to be accessed locally or remotely, which makes information sharing easier, especially for companies that have multiple factories located all over the world. CMMS deployment can seriously contribute towards improving organizational performance by ensuring improved communication and decision-making capabilities within the maintenance function. Accessibility of information and communication links on CMMS provide improved communication of repair needs and work priorities, improved coordination through closer working relationships between maintenance with production and engineering, and increased maintenance responsiveness.

17.6.1.8 Visual Workplace

The visual workplace is not merely a system for keeping track of tools and equipment. It is a program that boosts organizational performance and supports business objectives of lowering costs by operating with minimum waste. Visual management (or visual communication) must be holistically deployed to boost the company's productivity through increasing effectiveness of employees by effective sharing of information and encouraging workers to participate in developing this information. The visual workplace management comprises of visual controls and visual systems (Table 17. 2).

Visual Controls are simple signals that provide an immediate understanding of a situation or condition. They are very efficient, self-regulating, and worker managed. A visual workplace should include hundreds of visual control devices, where a visual device is: a mechanism or gadget which is intentionally designed to influence, direct, or limit behavior by making information vital to the task-at-hand available at-a-glance without speaking a word. The main purpose of visual controls is to organize a working area such that people (even outsiders) can tell whether things are going well or not without the help of an expert.

Table 17.2. Visual systems description in various locations at workplace

On the equipment	• Marking the proper operating ranges on temperature, pressure, flow and speed gauges • Temperature sensing tape on motors, reduction gears to check for overheating • Labeling lubrication and fluid fill points • Transparent covers for panels, belts to show direction • Marking directions of flow, feed or rotation to prevent installation errors • Using colour-coded grease fitting caps to protect and designate lubrication types and frequency • Floats with flower or doll mechanisms to show the flow of cooling water, coolant or any other fluid • Permanently attaching vibration analysis pickup discs to equipment and applying identification labels for reliable and repeatable vibration monitoring • Red and green zones on gauges, meters • Stickers for inspection (ear, hand) • Acrylic sheets to expose closed drives, transparent doors wherever possible • Labeling replacement belt, filter, chain sizes and part numbers on the equipment to save time looking up replacement part numbers • Color-coding set-up and changeover parts for specific product sizes • Using problem tags to pinpoint the location of machine problems and to request maintenance using a visual ‘action board’ • Labeling pneumatic lines and devices to aid troubleshooting • Labeling electrical and electronic wiring and devices to aid troubleshooting • Match-marking nuts and bolts to visually indicate proper tightness • Labeling inspection points and gauge reading sequence numbers
--------------------------------	---

Table 17.2. (continued)

In the spare parts room	• Inventory control cards with photograph of parts, part numbers, lead time for re-ordering, supplier or source, minimum/maximum levels. • Reorder signal cards placed at the minimum inventory level
In the area near the equipment	• Equipment action boards in the plant communicate performance trends and improvements • Paper fans near motor to see the cooling unit operation. • Pressure drop by toys based on unfurling/uncoiling of toys • Visual preventive maintenance (PM) schedules showing when PMs are due, past due and completed for the entire year • Equipment loss structure and list of Improvement projects with status • Any work instruction, to be conveyed to the operator who is not present in that shift • Display of inspection route map • Photographs and small drawings to show important points in procedures • Photographs to show where to inspect or adjust
Visual procedures and work instructions	• Photographs and small drawings used to show important points in procedures • Photographs used to show where to inspect or adjust • Photographs used to show where to get equipment readings for a shift inspection log sheet

Visual systems must be effectively deployed in the workplace to put an end to the biggest enemy of our workplace, *i.e.*, motion (delay due to unnecessary movement), by visually sharing information vital to the task-at-hand with the people who need it the most – operators, supervisors and managers, without speaking a word, *i.e.*, visually, so that the work can be performed with clarity, precision and confidence. It includes visual order, visual measures and visual standards. Visual order is a methodology with two important purposes. The purpose is to help in preparing physical workplace to hold visual location information and visual specification information through borders, home addresses and (if possible) ID labels. Visual measures are means for measuring overall manufacturing performance through visual displays. Visual standards help in monitoring it.

17.6.2 TPM Initiatives Implementation Phase (Phase II)

The detailed description of various issues related to TPM implementation initiatives has been elaborated for strategically focusing upon different aspects of TPM implementation. The various issues related to holistic TPM implementation initiatives to be holistically followed in the organization has been elaborated here for realizing true potential of TPM.

17.6.2.1 *Autonomous Maintenance Initiatives*

The manufacturing organization should encourage equipment operators to work alongside maintenance workers, as part of the TPM program, to perform tasks that prevent deterioration of production equipment. In TPM, this type of operator involvement in maintenance activities is called autonomous maintenance (AM). The organization needs to recognize that equipment operators have significant potential for making contributions to improvement in equipment performance, since the 'I run it, you fix it' attitude cannot effectively eliminate breakdowns and defects. The organization should endeavor to build the sense of ownership of the equipment and adapt autonomous maintenance initiatives through proactive involvement of equipment operators to eliminate thoroughly failures, stoppages, and defects and accelerated equipment deterioration. The organization should train the operators to perform autonomously routine cleaning, lubrication, tightening, adjustment, inspection, and re-adjustment (C-L-T-A-I-R). Organization should work towards developing operator proficiency skills in equipment mechanisms. Further, operators must work to develop a deeper understanding of their equipment which should improve their operating skills. The operators should be encouraged to get involved with routine maintenance and improvement activities that halt accelerated deterioration, control contamination, and help obviate equipment problems.

The basic objectives of an autonomous maintenance program should include: addressing accelerated deterioration (cleaning the equipment, identifying and correcting problematic areas, developing and implementing standards); bringing equipment back to normal (affecting equipment improvements, identifying and correcting chronic problems and hard to access areas, developing and implementing standards); and further improving the equipment (knowing the mechanics of production systems, innovating and implementing new ideas). Autonomous maintenance practiced by an operator or manufacturing work cell team member helps to maintain high machine reliability, ensure low operating costs, and maintain high quality of production parts.

17.6.2.2 *Focussed Improvement Initiatives*

The organization must adopt focussed improvement initiatives (Kobetsu Kaizen-KK) to maximize efficiency by eliminating wastes and manufacturing losses (Figure 17.10). The losses to be addressed through strategic KK initiatives must include seven major losses impeding equipment efficiency (failure losses, setup/adjustment losses, reduced speed losses, idling/minor stoppage losses, defect/rework losses, startup losses and, tool changeover losses), losses impeding machine loading time (planned shutdown losses), five major losses impeding

human performance (distribution/logistic losses, line organization losses, measurement/adjustment losses, management losses, and motion related losses) and three major losses impeding effective use of production resources (yield losses, consumable – jig/tool/die losses, and energy losses). These losses are structured to easily identify and agree on the opportunities for improvement. The effective performance measures like OEE must be introduced and managed for effectively implementing TPM. Improvement teams must be focused and management must identify opportunities and priorities for improvement teams, maximizing the limited resources available. As the team matures and becomes aligned to company objectives and expectations, then and only then should this decision making be delegated to the team.

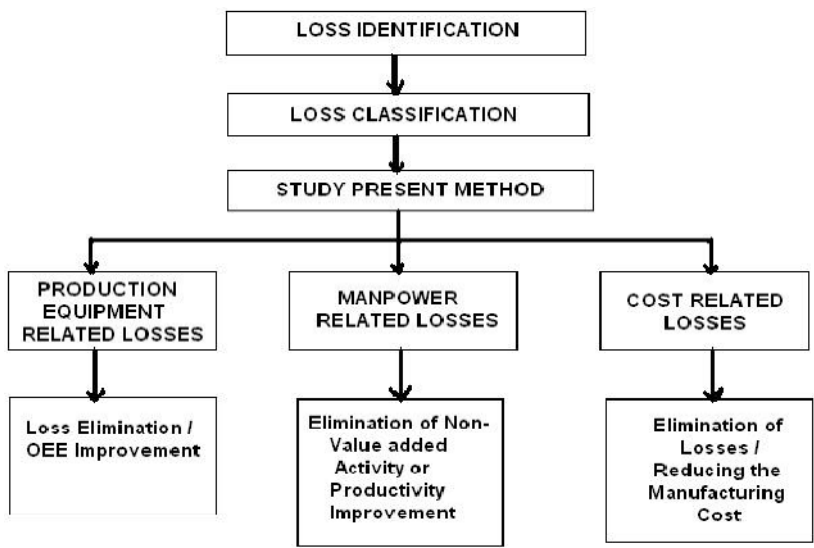


Figure 17.10. Issues involved in focused maintenance

17.6.2.3 Planned Maintenance Initiatives

It has been observed that, for successful TPM implementation, organization must harness competencies for improving traditional maintenance performance in the workplace. Through strategic planned maintenance initiatives, organization should endeavor to realize a condition of ‘zero failure’ in the plant at minimum possible maintenance cost by focusing on actions related to maintainability and reliability of production equipment, correcting equipment design weaknesses, and maintaining ideal state of equipment. In this regard, organization needs to develop standard work practices and safe operating procedures covering the entire range of production systems and also needs to ensure holistic implementation of laid out procedures by a motivated and competent workforce. The organization must impress upon addressing problems related to production systems by focusing on root causes of the problems, rather than emphasizing mere restorations.

The organization should endeavor to develop preventive maintenance programs to transform the existing maintenance schedule and system into an improved

maintenance process. PM should focus on various activities like daily maintenance to prevent deteriorations, periodic inspections or equipment diagnoses to measure deteriorations, and restoration to correct, recover from deterioration, thereby leading to realization of improved inspection processes, plans and schedules. While performing preventive maintenance, data must be collected for equipment effectiveness measurements, reliability studies, maintainability metrics, and operating costs.

The organization needs to adapt proactive maintenance initiatives by deploying corrective actions aimed at addressing ‘sources of failure’ to extend the life of mechanical machinery. Proactive maintenance provides a logical culmination to other types of maintenance, that is, reactive, preventive, and predictive maintenance. Figure 17.11 depicts proactive maintenance initiatives for eliminating recurrence of breakdowns. The organization must deploy various proactive maintenance strategies like failed part analysis, root cause failure analysis, reliability engineering, rebuild certification/verification, age exploration, and recurrence control to extend equipment life.. Further, information technology and CMMS based maintenance systems can also be typically used for automatic reporting and instant e-mailing or cell phone notification of critical events. The switch to proactive programs delivers more cost benefits (tenfold savings) than past journeys from ‘breakdown’ to ‘preventive/predictive’ strategies. With proactive maintenance, failure avoidance and a world-class maintenance program can truly be achieved.

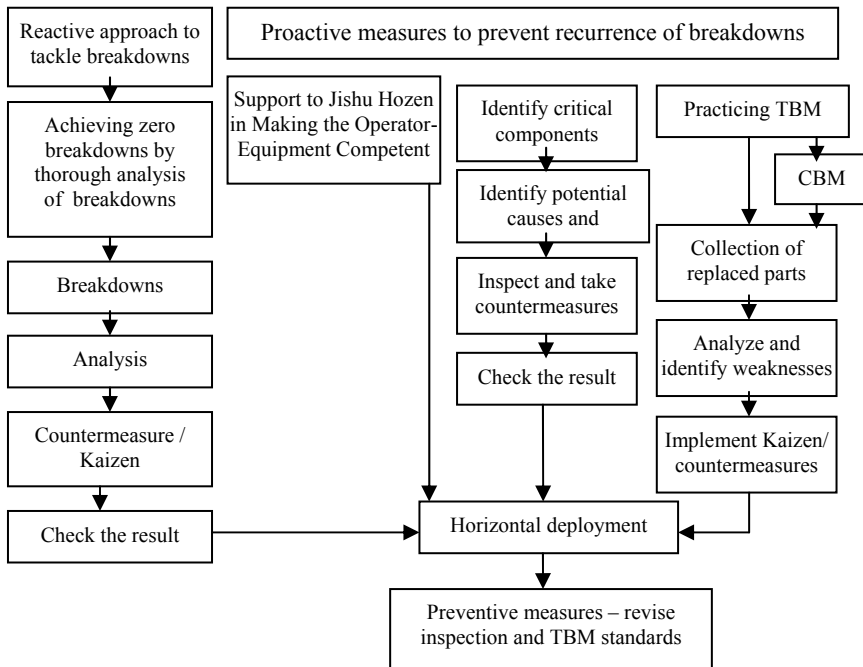


Figure 17.11. Proactive maintenance for preventing recurrence of breakdowns

17.6.2.4 Quality Maintenance Initiatives

The organization should endeavor to realize ‘zero defects’ and ‘zero customer complaints’ by supporting and maintaining equipment conditions through strategic quality maintenance (QM) initiatives. QM initiatives in the organization must involve various activities including overcoming deficiencies in quality system to achieve defect free product, setting conditions for zero defects, maintaining optimal machine and tooling conditions, maintaining equipment operation performance within standard ranges, inspecting and measuring conditions in time series, preventing occurrence of defects by periodic measurements and verification of standards, predicting possibility of quality defects by reviewing measured values, and taking counter measures in advance. The master plan for QM should involve data collection on defects for improving conditions to sustain zero defect conditions. QM initiatives should deploy various quality improvement techniques and strategies for achieving zero defects. The strategic tools and techniques recommended for meeting aforesaid objectives through QM program should include QC process diagram, process capability investigation chart, scatter diagram, X – R chart, QA matrix, defect phenomenon check sheet, pareto diagram, mechanism/function diagram, work standard sheet, 4M conditions survey chart, defect cause analysis, why – why analysis sheet, process point analysis chart, PM analysis sheet, improvement sheet, Ishikawa diagram, input output analysis, loss tree analysis, flow chart analysis, histogram analysis and Poka Yoke, *etc.*

17.6.2.5 Office TPM Initiatives

The key objectives of office TPM (OTPM) pillar should include realizing zero functional loss, organizing high efficiency offices and rendering service, and support to production departments by focusing on effective workplace organization and standardized work procedures. The various initiatives recommended for organization for realizing aforesaid objectives include: autonomous maintenance activities in administrative and other indirect departments; individual improvement (Kaizen) activities in administrative and other indirect departments; and providing administrative support for production departments. The responsibilities assigned for improving organizational performance through strategic OTPM initiatives include: providing and maintaining clean, bright, hazard-free, safe and pleasant working environments; ensuring good house-keeping; eliminating procedural and process delays/losses; increasing manpower productivity; eliminating deterrents in smooth production of goods; optimally managing stores and spares inventory; optimizing investigation time for mitigation of customers complaints; reducing delays in loading and unloading of materials; addressing complaints related to errors in payment of wages; appropriate office management; addressing delays in payment of supplier bills; ensuring procurement of raw materials and supplies at optimal quality and cost; reduction in administrative expenses; and optimizing manufacturing cycle time. Strategic OTPM initiatives offer efficient and cost effective procedures in the organization to improve efficiency, work organization, housekeeping, and quality.

17.6.2.6 Safety, Health and Environment Initiatives

Safety, health, and environment pillar facilitates the organization in achieving standard operating practices in the workplace, safe working environment; motivated employees; and pollution free, clean and green environment (Figure 17.12). Though this pillar appears later in the TPM roadmap, it acts like a real eye opener and helps in affecting cultural transformation faster than other pillars. It is the pillar which make shop-floor people understand that they are an important part of the organization. This instiles improved confidence among all the employees and they understand that TPM initiatives can effectively contribute towards employee's wellbeing and safety. The necessary steps must be taken to eliminate unsafe practices and conditions like: missing or broken belts, chains, coupling guards, hand railings, gratings, drain covers, sharp corners; missing emergency stop devices; uneven or oils floors; inadequate lighting; congested places; and sources of fumes, gases, heat, or vibrations. Adequate training regarding safety should be imparted and awareness created amongst all employees against careless working attitudes and the employees should be motivated to follow safety norms. The employees should be provided with adequate safety gadgets like safety helmets, nose masks, safety belts, safety shoes, hand gloves, safety goggles, safety clothing, insulated tools and other related personnel protective equipment as demanded by the hazardous/dangerous situations. The safety promotional activities like: best safety Kaizen, safety slogan competition; safety poster, essay, speech, poem competition; safety month, week celebration; and reporting best near-miss initiatives can contribute highly in improving safety in the workplace. The organization should endeavor to tackle abnormal working conditions and ensuring a safe, hygienic working environment to all concerned in the organization.

17.6.2.7 Development Management Initiatives

Development management initiatives facilitate the organization to reduce dramatically the time from initial development to full-scale production and achieve vertical startup through maintenance prevention (MP) and early product management in development of new products. This component of TPM is responsible for incorporating the knowledge and manufacturing competencies gained from maintaining existing equipment into new equipment designs. This information includes equipment performance, life cycle costs, reliability and maintainability targets, equipment testing plans, operating documentation, and training.

The organization needs to adapt key maintenance prevention (MP) initiatives at the new equipment introduction stage, to ensure that new equipment is safe, easy to use and maintain, free from failures, and unlikely to produce defects. This process requires joint planning and coordinating with other stakeholders involved in equipment start-up for accomplishing rapid and reliable ramp-up to designed production rate performance. The concerted efforts should be made for affecting manufacturing system performance improvements by emphasizing maintenance prevention initiatives and enhancing focused production system improvements by fostering competencies related to production facilities by deploying feedback from customers and various departments, focusing upon learning from existing equipments to new systems, incorporating design related improvements, improving

safety at workplace, and integrating TPM with other performance improvement initiatives.

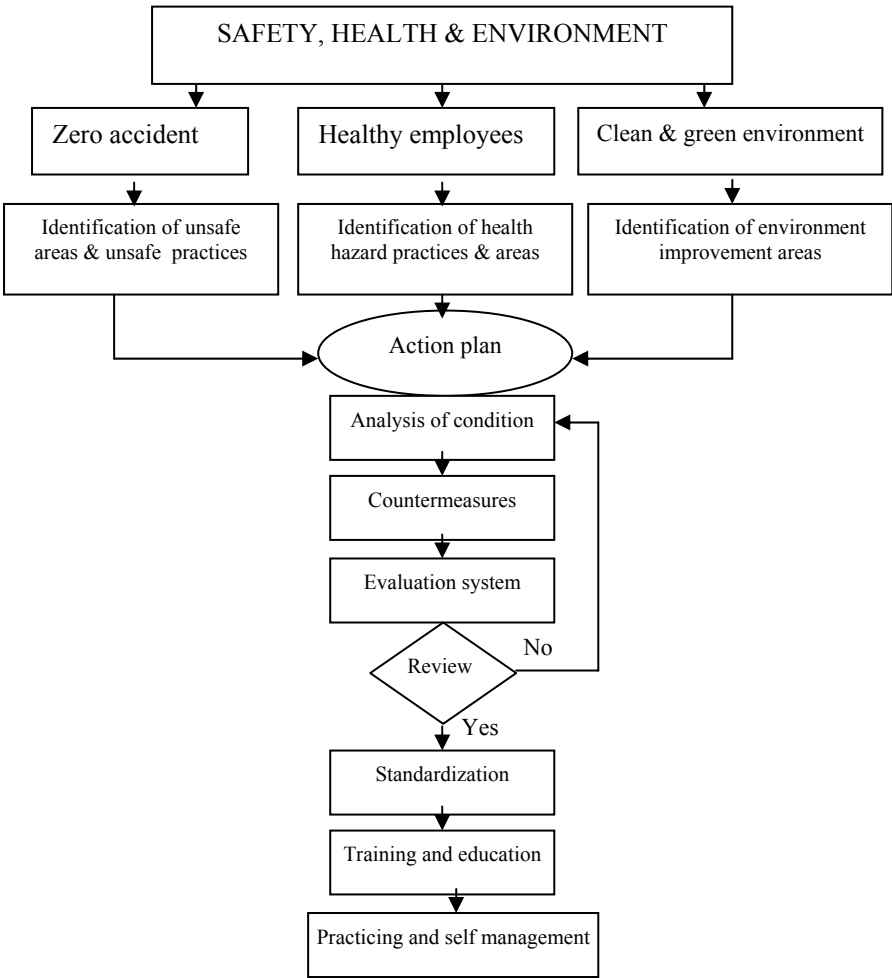


Figure 17.12. Strategies for improving safety, health and environment

17.6.2.8 Tool Management Initiatives

The scope of tool management initiatives is to eliminate machine stoppages due to tools (eliminating downtime due to non-availability of tools, reducing downtime due to poor quality of tools, reducing downtime due to tool reset) and reducing tool consumption cost (increasing life, reducing price). The basic conditions must be provided in the workplace by practicing 1S and 2S improvements for production and maintenance tools in tool stores, providing appropriate visualization for tool-search in tool store, practicing 1S and 2S improvements at tool room and tool boxes at the production lines. The situations contributing to machine stoppages due to poor tool quality should be identified and appropriately addressed for ensuring

effective machine utilization. The situations amounting to poor tool quality can be classified as: blunt tool/breakage (wrong grinding, poor geometry, poor tool material, poor machining condition, and poor component material) and first piece defect (inspection error, design error, absence of re-grinding and re-sharpening procedures). Appropriate measures must be taken to address machine stoppages due to tooling problems. The tool replacement procedures, adjustment procedures and design mechanisms should be developed for eliminating problems due to tool resetting.

17.6.2.9 Maintenance Benchmarking Initiatives

The maintenance benchmarking study provides an organization with an opportunity to compare the prevailing maintenance practices and performance with selected best-of-the-best practices, identifying strengths and improvement opportunities across various best practice areas of maintenance management and establishing a sound foundation of quantitative and qualitative strategy for comprehensive maintenance performance improvement. A comprehensive evaluation of maintenance and reliability practices across the ‘best practice’ areas of leadership, human resources, planning and scheduling, preventive and predictive maintenance, reliability, materials management and contract maintenance management can be undertaken for achieving significant improvements in maintenance function performance.

The organization can ultimate focus upon following and realizing “world class maintenance performance benchmark indices” for evaluating the success of TPM implementation programs. Various “world class maintenance performance benchmark indices” include: planned maintenance work (90%); schedule compliance (70%); work order discipline (> 90%); process availability (> 95%); quality rate (99%); speed rate (90%); OEE (85%); maintenance cost as a percentage of total sales (< 3%); maintenance cost as % of RAV (< 2%); labor utilization (wrench time) (50–60%); maintenance overtime (6%); percentage of labor to materials cost (70%:30%); maintenance callback (3%; target 0); and MTBF for pumps (7 years).

The benchmarking study provides an understanding of maintenance performance based on a comprehensive range of financial, personnel and management comparison parameters. This enables a sound understanding to be developed of comparative maintenance performance. The key issues and improvement opportunities are identified and a dollar business stake is estimated showing direct benefits to the organization by achieving maintenance excellence.

17.6.3 Standardization Phase (Phase III)

Finally, TPM initiatives need to be stabilized and holistically pursued over a reasonable period of time for reaping true potential from TPM implementation. The manufacturing organization should continue to explore and enhance the various TPM themes being pursued. TPM initiatives should be horizontally deployed to all organizational activity areas/departments besides maintenance and manufacturing functions, so as to include: R & D, design, product development, service areas, assembly work, procurement, sales, marketing, administrative,

management (accounting, general affairs, planning and quality assurance), and production scheduling. The critical aims of deploying TPM initiatives to non-production departments is to improve effectiveness of departments in fulfilling their functions, as well as, improving effectiveness of support rendered by these departments to production activities.

17.6.3.1 Deploying Key Performance Indicators 'KPI' for Assessing Manufacturing Performance

Performance metrics are essential within the physical asset management process, as they help management and plant personnel to understand business and mission requirements and identify opportunities to increase effectiveness and measure performance to objectives. The organization must deploy KPIs to measure specific parameters across all classes of metrics. Equipment management metrics must be concise and connect directly to corporate/mission objectives and demonstrate contributions to manufacturing effectiveness. Focusing on too many areas at once may result in information overload and increase difficulty in directing limited resources to highest value activities. KPIs are necessary to establish objectives, measure performance and, reinforce positive behaviors for realizing World Class Maintenance.

17.6.3.2 Deploy Lean Manufacturing Practices

The enhanced application of TPM tools and practices to cater to an organization's overall growth and sustainability endeavors calls for deploying lean manufacturing practices to satisfy ever growing organizational demands of the global organizations. Various lean manufacturing practices reportedly effectively deployed by global organizations for attaining global leadership, along with typical TPM programs, include JIT manufacturing, TQM, 6- σ , benchmarking, and quality function deployment. There are many success stories and research reported on TQM, JIT and TPM. It is therefore strongly recommended that TQM, JIT programs be strategically adopted with the TPM programs to garner overall manufacturing and organizational competencies. Further, manufacturing organization can also deploy other lean manufacturing strategies like Continuous flow manufacturing, cellular manufacturing, benchmarking, levelled manufacturing and reverse engineering for still greater organizational performance enhancement. The organization must heed the warning that using too many strategies simultaneously at an early stage may lead to confusions and dilution of the perceived impact of an individual strategy. However, manufacturing organizations can deploy different proactive lean manufacturing strategies during the stabilization stage of TPM implementation process to enhance further the organizational performance.

17.6.3.3 Sustain TPM Initiatives

Finally, concerted efforts must be made to ensure sustained TPM deployment in the manufacturing organization, as manufacturing improvements are only possible through persistent deployment of world class TPM initiatives. The goal of organization at this stage, after successful deployment of TPM, has to continue the TPM program into the incremental process improvement phase, using a continuous

quality improvement (CQI) approach. It is extremely important for an organization to move consistently forward after attaining TPM excellence award for sustaining achievements realized and to achieve higher levels of performance. The changes introduced into the organization through strategic TPM activities must be anchored, thereby becoming an established part of everybody's daily routine. TPM has to be regarded as a 'change process', rather than a 'project', otherwise the competencies gained by the organization might fade away after the project is completed. In order to retain all the productivity gains made through successful TPM implementation, the organization needs to create an organizational infrastructure to sustain all the new TPM behaviors. The challenge to implementing a sustainable TPM process is understanding whether the TPM process is being implemented correctly and knowing where the weaknesses are. Thus, a TPM audit process and TPM gap analysis must be put into place for evaluating the evolution of permanent changes taking place in the organization. The appropriate auditing and monitoring system should be developed to improve TPM results continuously. The sustained TPM programs have the capability to achieve "world class organization" and to assume leadership roles in the competitive environments.

17.7 Barriers in TPM Implementation

TPM implementation is not an easy task by any means. The number of organizations successfully implementing TPM program is considered relatively small. While there are several success stories and research on TPM, there are also documented cases of failures in implementation of TPM programs in different situations. TPM demands not only commitment, but also structure and direction. The prominent problems in TPM implementation include cultural resistance to change, partial implementation of TPM, overly optimistic expectations, lack of a well defined routine for attaining the objectives of implementation (equipment effectiveness), lack of training and education of TPM teams on 'whats and whys of TPM', failure to start with operator-involved maintenance, superficial TPM deployment, ineffective rewards and felicitation mechanisms, lack of organizational communication, and implementation of TPM to conform to societal norms rather than for its instrumentality to achieve world class manufacturing.

The various obstacles hindering an organization's quest for achieving excellence through TPM initiatives have been classified as organizational, cultural, behavioral, technological, operational, financial, and departmental barriers (Ahuja and Khamba, 2008b).

The organizational obstacles affecting successful TPM implementation in organizations include:

- Organization's inability to bring about cultural transformations;
- Organization's inability to implement holistically change management initiatives;
- Lack of commitment from top management and communication regarding TPM;

- Lack of understanding of TPM concepts and principles;
- Inability of management to educate stubborn employee unions about true potential of TPM;
- Organization's inability to change mindset of workforce to obtain total employee involvement;
- Wrong pace of TPM implementation and focusing on too many improvement initiatives;
- Inadequacies of reward and recognition mechanisms in the organizations;
- Inadequacies of master plan in the absence of a focused approach;
- Middle management's resistance towards offering empowerment and recognition of bottom level operators due to fears of loss of authority and respect;
- Inability to adhere strictly to laid out TPM practices and standards;
- Organization's inability to enhance employee competencies towards job;
- Alienation of employees from growth and sustainability endeavors of organizations;
- Lack of awareness of TPM concepts and principles among the employees;
- Inadequate services for the employees in organizations; and
- Absence of mechanisms to critically evaluate and monitor maintenance performance metrics like overall equipment effectiveness (OEE), return on net assets (RONA) and return on capital employed (ROCE).

The cultural obstacles affecting successful TPM implementation in organizations include:

- Inability to align employees to organizational goals and objectives;
- Lack of professionalism including lack of consistency, resistance to change, poor quality consciousness coming in the way of organizational transformations;
- Strong unions, rigid mindsets, non-flexible approaches, non-adaptable attitudes;
- Stubborn attitudes regarding existing organization, knowledge and beliefs;
- Inability of top management to motivate employees to 'unlearn to learn';
- Concern of employees with 'what's in it for me' attitude;
- Low skill-base also a deterrent to accept changes in the workplace;
- Marginal employee participation in organizations towards decision making; and
- Compromising attitude on quality of production with rework accepted as part of production activities.

The behavioral obstacles affecting successful TPM implementation in organizations include:

- Resistance from employees to adapt to proactive, innovative management concepts;
- Occasional difficulties to succeed as cross functional teams (CFT);
- Lack of motivation on part of employees to contribute effectively towards organizational development and sustainability efforts;
- Functional orientation and loyalty;

- Inadequate efforts towards multi-skilling and periodic skill updation of employees;
- Lack of willingness on part of operators to learn more regarding functioning of production systems; and
- Resistance to accept changes due to job insecurity and apprehension of loss of specialization due to technological improvements.

The technological obstacles affecting successful TPM implementation in organizations include:

- Little emphasis to improve production capabilities beyond the design capabilities;
- Inadequate initiatives to assess and improve reliability of production systems and ensure the faster, dependable deliveries;
- Highly inadequate predictive maintenance (Pd.M.) infrastructural facilities in the organizations;
- Highly inadequate computerized maintenance management systems (CMMS) infrastructural facilities in the organizations;
- Absence of mechanisms for investigating inefficiencies of production system (losses, wastes) leading to lack of impetus for affecting manufacturing improvements;
- Poor flexibilities offered by production systems due to long set up and changeover times;
- Less educated workforce due to inadequacies of training on emerging technologies;
- Lack of training opportunities and skills regarding quality improvement techniques and problem diagnostics;
- Little emphasis on maintenance prevention initiatives regarding possibilities of improvements in existing products and manufacturing systems; and
- Poor energy efficiency of production systems.

The operational obstacles affecting successful TPM implementation in organizations include:

- General acceptance of reasonably high levels of defects associated with production systems with little emphasis on realization of world-class six-sigma production capabilities;
- Non-adherence to standard operating procedures (SOP);
- Little empowerment to operators to take equipment related or improvement decisions;
- Absence of planned maintenance (PM) check-sheets to conduct routine maintenance jobs efficiently;
- Apathy of top management to implement safe work practices at the workplace;
- Resistance from production operators to perform basic autonomous maintenance tasks;
- Poor and non-encouraging workplace environments in the absence of 5S implementation;

- Little motivation or time available for affecting process related improvements, while major focus of organizations is on meeting routine production targets by all means; and
- Emphasis on restoration of equipment conditions rather than prevention of failures.

The financial obstacles affecting successful TPM implementation in organizations include:

- Requirement of significant additional resources in the beginning of TPM implementation program with moderate performance improvements in initial stages of TPM;
- Inability of top management to support improvement initiatives due to resource crunch; and
- Absence of appropriate motivating reward and recognition mechanisms.

The departmental obstacles affecting successful TPM implementation in organizations include:

- Low synergy and coordination between maintenance and production departments;
- Reluctance of production operators to accept autonomous maintenance initiatives as part of their routine jobs;
- Firm divisions between maintenance and production function responsibilities; and
- A general lack of trust by maintenance department in productive operator's capabilities for performing basic autonomous maintenance tasks.

Thus, it can be asserted that there are many factors that may contribute to the failure of the organizations to implement TPM successfully and reap the true potential of TPM. TPM implementation requires a long-term commitment to achieve the benefits of improved equipment effectiveness. Training, management support, and teamwork are essential for the success of TPM implementation programs. Thus, it becomes pertinent to develop TPM support practices like committed leadership, vision, strategic planning, cross-functional training, employee involvement, cultural changes in the organizations, continuous improvement, motivation, and evolving work related incentive mechanisms in the organizations to facilitate TPM implementation programs to realize world class manufacturing attributes.

17.8 Success Factors for Effective TPM Implementation

TPM is a result of the corporate focus on making better use of available resources. There are many success criteria for effective and systematic TPM implementation. In order to realize the true potential of TPM and ensure successful TPM implementation, TPM goals and objectives need to be fully integrated into strategic and business plans of the organizations, because TPM affects the entire organization and is not limited to production. The first course of action is to establish strategic directions for TPM. The transition from a traditional

maintenance program to TPM requires a significant shift in the way production and maintenance functions operate. Rather than a set of instructions, TPM is a philosophy, the adoption of which requires a change of attitude by production and maintenance personnel. The key components for successful implementation of TPM have been envisioned as worker training, operator involvement, cross functional teams, and preventive maintenance. There is an utmost need to foster initiatives facilitating smooth TPM implementation that include committed leadership, strategic planning, cross-functional training, and employee involvement. In order to capture the TPM program completely, it is pertinent to combine TPM practices identified as pillars or elements of TPM with the TPM development activities. For TPM to be successful, the improvement initiatives must be focused on benefiting both organization and employees. There is a need to foster an environment for facilitating employees to adapt and implement smoothly the autonomous maintenance and planned maintenance postulates of TPM implementation.

There is an urgent need for establishing and holistically adopting key enablers and success factors in the organizations to ensure success of the TPM implementation program by harnessing total participation of all employees in the organizations. The key enablers and success factors for successful implementation of TPM has been classified into six categories:

1. Top management contributions;
2. Cultural transformations;
3. Employee involvement;
4. Traditional and proactive maintenance policies;
5. Training and education; and
6. Maintenance prevention and focused production system improvements.

The strategic issues related to various TPM enablers and success factors have been explained using an Ishikawa diagram (Ahuja and Khamba, 2008b). It is strongly believed that holistic adaptation of enablers and success factors can obviate the ill effects of obstacles to TPM implementation and can strategically lead the organizations to harness manufacturing competencies for sustained competitiveness.

Thus, organizations need to develop an understanding of restraining and driving forces and need to take proactive initiatives for overcoming the hindrances caused by the obstacles to successful TPM implementation programs for reaping the true potential of TPM. Only steadfast adherence to the TPM vision and well chalked out master plans can effectively lead to success of TPM implementation programs. Thus, organizations need to accept in the true spirit that TPM is implemented right the first time, even if it takes a little longer. The organizations must realize that shortcuts and unrealistic schedules and over-aggressive plans might result in failures, restarts, and loss of motivation to implement TPM consistently over a long period of time. The key is to learn from mistakes and make subsequent efforts better.

17.9 Summary

Holistic TPM implementation can lead to the establishment of strategic proactive maintenance practices in the organization for avoiding future system and equipment related losses and marshal the organizations towards capability building for sustained competitiveness. TPM is not a radically new idea; it is simply the next step in the evolution of good maintenance practices. TPM is indispensable to sustain just-in-time operations. TPM facilitates immensely the organizations in improving the synergy between maintenance department and rest of the production functions, resulting in eliminating defects, improving manufacturing process reliability, improving overall equipment effectiveness, and reducing costs, thereby affecting sustainability efforts of the organization to meet cut-throat global competition for business excellence. TPM has proved to be a means to supplement the concerted improvement efforts by addressing equipment and other related problems that adversely affect the performance of the manufacturing system. Thus, in a highly competitive scenario, TPM can prove to be the best proactive strategic initiative that can lead organizations to scale new levels of achievements and could really make the difference between success and failure of organizations.

TPM implementation in an organization can contribute effectively in realization of world class manufacturing. However, it must be understood that a TPM implementation program does not yield overnight success and it requires a reasonable period of holistic interventions, varying between 3 and 5 years, to realize the true potential of TPM. It takes appropriate planning and a focused TPM implementation plan, adequately assisted by top management through imbining organizational cultural improvement over a considerable period of time, to realize significant manufacturing performance improvements from the holistic TPM implementation program. Thus it can be concluded that for the successful implementation of a TPM program in the organization, it becomes mandatory for the manufacturing managers to understand the functioning and interaction of the different facets of TPM, so that the concept can fulfill its true potential.

References

- Ahuja IPS and Khamba JS (2007) An Evaluation of TPM Implementation Initiatives in an Indian Manufacturing Enterprise. *J of Qual in Maint Eng* 13(4): 338–352.
- Ahuja IPS and Khamba JS (2008c) An Evaluation of TPM Initiatives in Indian Industry for Enhanced Manufacturing Performance. *Intl J of Qual and Reliab Manag* 25(2): 147–172.
- Ahuja IPS and Khamba JS (2008a) Total Productive Maintenance Implementation in a Manufacturing Organization. *Int J of Prod and Qual Manag* 3(3): 360–381.
- Ahuja IPS and Khamba JS (2008b) Strategies and Success factors for overcoming Challenges in TPM Implementation in Indian Manufacturing Industry. *J of Qual in Maint Eng* 4(2).
- Bamber CJ, Sharp JM and Hides MT (1999) Factors affecting successful implementation of Total Productive Maintenance: A UK Manufacturing Case Study Perspective. *J of Qual in Maint Eng* 5(3): 162–181.
- Bonavia T and Marin JA (2006) An empirical study of lean production in the ceramic tile industry in Spain. *Int J of Oper and Prod Manag* 26(5): 505–531.

- Cua KO, McKone KE and Schroeder RG (2001) Relationships between implementation of TQM, JIT and TPM and manufacturing performance. *J of Oper Manag* 19(6): 675–694.
- Cua KO, McKone KE and Schroeder RG (2006) Improving Performance through an Integrated Manufacturing Program. *The Qual Manag J* 13(3): 45–60.
- Gregory A (2006) Number cruncher – overall equipment effectiveness and total productive maintenance. *Work Manag* 59(7): 18–20.
- Heston T (2006) Culture change for maintenance. *Fabr and metalwork* 5(9): 70–72.
- Lycke L and Akersten PA (2000) Experiences of implementing TPM in Swedish industries. *Int J of Reliab and Appl* 1(1): 1–14.
- McCarthy D (2004) *Lean TPM*, Butterworth Heinemann, U.K.
- McKellen C (2005) Total Productive Maintenance. *MWP* 149(4): 18.
- McKone KE, Roger GS and Cua KO (1999) Total productive maintenance: a contextual view. *J of Oper Manag* 17(2): 123–144.
- Naguib H (1993) A Roadmap for the Implementation of Total Productive Maintenance (TPM) in a Semiconductor Manufacturing Operations, Semiconductor Manufacturing Science Symposium, ISMSS–1993., IEEE/SEMI International, 19–20 July: 89–97.
- Nakajima S (1988) *Introduction to TPM*, Productivity Press Inc., Cambridge, MA.
- Nakajima S (1989) *TPM Development Program: Implementing Total Productive Maintenance*, Productivity Press Inc., Cambridge.
- Pirsig RM (1996) Total productive Maintenance, in ‘Managing Factory Maintenance’ by Levitt, J., Industrial Press Inc., Medison Avenue, New York.
- Rhyne DM (1990) Total plant performance advantages through total productive maintenance, Conference Proceedings, APICS, Birmingham: 683–686.
- Robinson CJ and Ginder AP (1995) *Implementing TPM: The North American Experience*, Productivity Press Inc., Portland, OR, New York.
- Rodrigues M and Hatakeyama K (2006) Analysis of the fall of TPM in companies. *J of Mate Process Technol* 179(1–3): 276–279.
- Shirose K (1996) *Total Productive Maintenance: New Implementation Program in Fabrication and Assembly Industries*, Japan Institute of Plant Maintenance, Tokyo, Japan.
- Steinbacher HR and Steinbacher NL (1993) *TPM for America*, Productivity Press, Portland, Oregon.
- Wang FK (2006) Evaluating the Efficiency of Implementing Total Productive Maintenance. *Total Qual Manag* 17(5): 655–667.
- Willmott P (1994) Total Quality with Teeth. *The TQM Magazine*, 6(4): 48–50.
- Wireman T (1990) *Total Productive Maintenance – An American Approach*, Industrial Press Inc., New York.
- Witt CE (2006) TPM: The Foundation of Lean. *Mater Handl Manag* 61(8): 42–45.

Warranty and Maintenance

D.N.P. Murthy and N. Jack

18.1 Introduction

Businesses use equipment to deliver their outputs (products and/or services) and individuals use consumer products to satisfy their personal needs and provide entertainment. Both types of item are getting more complex due to rapid advances in technology and to the increasing expectations of customers (businesses and individuals). These customers need to be assured that the items will perform satisfactorily over their useful lives and one way of providing this assurance is through a product warranty. Most countries have either enacted or are in the process of enacting stricter warranty legislation to protect customers' interests. Manufacturers are required to rectify all failures that occur over the warranty period and this is referred to as warranty servicing. Warranty servicing results in additional costs to the manufacturer and it has been reported in the literature that these costs can vary between 2% and 10% of an item's sale price depending on the product and manufacturer.

In warranty servicing, corrective maintenance (CM) actions are performed to restore a failed item to an operational state and preventive maintenance (PM) can also be used to reduce item degradation and the risk of failure. Maintenance is therefore important in the warranty context since it has a major impact on warranty servicing costs. This chapter deals with this topic, and we discuss the issues involved and review the relevant literature.

Section 18.2 outlines some of the concepts involved in maintenance modelling. In Section 18.3 we discuss different types of base warranty, warranty costs, and extended warranties. Section 18.4 describes the link between warranties and maintenance, and the issues concerned with maintenance logistics for warranty servicing are covered in Section 18.5. In Section 18.6 we deal with the outsourcing of warranty servicing and we state our conclusions and suggest topics for future research in Section 18.7.

18.2 Maintenance Modelling

In this section we give a brief overview of some basic concepts needed for maintenance modelling. We use the term “item” to indicate either a piece of business equipment or a consumer product.

18.2.1 Reliability

Every item degrades with age and usage and ultimately fails. Failures occur in an uncertain manner and are influenced by factors such as design, manufacture (or construction), maintenance, and operation. Item reliability conveys the concept of dependability or absence of failure and is defined as follows by Blischke and Murthy (2000):

“the *reliability* of an item is the probability that it will perform its intended function for a specified time period when operating under normal (or stated) environmental conditions.”

18.2.2 Types of Maintenance

PM actions control (or reduce) item degradation and produce an improvement in reliability. CM actions restore a failed item to an operational state and may either have no effect on reliability (minimal repairs) or may produce an improvement (imperfect repairs). Both CM and PM are discussed in the maintenance literature, which is extensive. It includes several review papers, and the most recent of these have been written by Cho and Parlar (1991), Dekker and Scarf (1998), and Wang (2002).

18.2.3 Failure Modelling

Item failures can be modelled using either one-dimensional (1-D) or two-dimensional (2-D) formulations and these are needed for 1-D and 2-D warranty analysis, respectively.

18.2.3.1 One-dimensional Model Formulations

Time to first item failure can be modelled using a distribution function $F(t)$ with density function $f(t)$ and hazard function $r(t)$. The modelling of subsequent failures depends on the quality of the CM actions and, when PM is used, this can affect both the first and subsequent failures. Many different distributions have been used to model item failure and these can be found in most books on reliability; see, for example, Meeker and Escobar (1998) and Blischke and Murthy (2000).

The concept of the rate of occurrence of failures (ROCOF) is used to model the failure behaviour of an item over time, allowing for the effect of PM and CM actions. The conditional ROCOF characterises the probability of item failure in the

interval $[t, t + \delta t)$ given $\mathfrak{T}(t)$, the history of failures and maintenance actions over the interval $[0, t)$. It is defined by

$$\lambda(t|\mathfrak{T}(t)) = \lim_{\delta t \rightarrow 0} \frac{P\{N(t + \delta t) - N(t) > 1 | \mathfrak{T}(t)\}}{\delta t} \quad (18.1)$$

where $N(t)$ is the number of failures in the interval $[0, t)$. Since the probability of two or more failures in the interval $[t, t + \delta t)$ is zero as $\delta t \rightarrow 0$, this conditional intensity function is equal to the derivative of the conditional expected number of failures, so

$$\lambda(t|\mathfrak{T}(t)) = \frac{d}{dt} E\{N(t) | \mathfrak{T}(t)\}. \quad (18.2)$$

For further discussion of this concept, see Ascher and Feingold (1984).

When an item is not subjected to any PM actions, all failures are repaired minimally (see Barlow and Hunter, 1960), and repair times are small relative to the time between failures, then the ROCOF has the same form as the hazard function of time to first failure, and so $\lambda(t|\mathfrak{T}(t)) = \lambda(t) = r(t)$.

Imperfect PM actions can be modelled by a reduction in (1) the failure intensity function, or (2) the item's virtual age.

Reduction in intensity function

Let $\lambda(t)$ and $\lambda_0(t)$ denote the intensity function (ROCOF) with and without any PM actions. The time to carry out any PM action is small relative to the mean time between failures and so can be ignored. The effect of PM on the intensity function is given by

$$\lambda(t_j^+) = \lambda(t_j^-) - \delta_j \quad (18.3)$$

where δ_j is the reduction resulting from the PM action at time t_j , $j \geq 1$. δ_j depends on the level of PM effort used and is constrained as follows:

$$0 \leq \delta_j \leq \lambda(t_j^-) - \lambda(0) \quad (18.4)$$

This ensures that PM actions cannot make the system better than new. The form of the intensity function is then

$$\lambda(t) = \lambda_0(t) - \sum_{i=0}^j \delta_i, \quad t_j < t < t_{j+1}, \quad (18.5)$$

for $j \geq 0$, with $t_0 = 0$ and $\delta_0 = 0$. Note that this implies that the reduction resulting from the PM action at t_j lasts for all $t \geq t_j$ as shown in Figure 18.1.

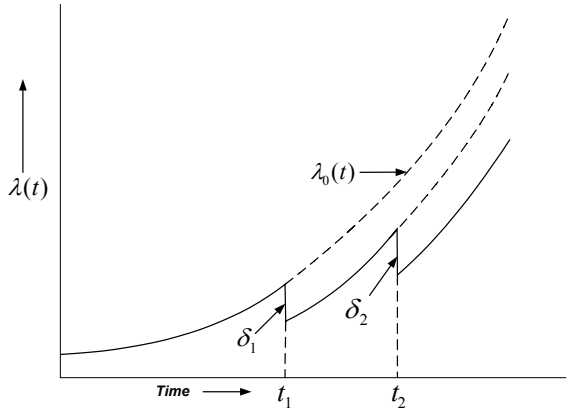


Figure 18.1. Effect of ‘intensity reduction’ PM actions

Reduction in age

A used item can be subjected to an upgrade (or overhaul) where components that have degraded significantly are replaced by new components so that the item becomes in a sense younger (from a reliability point of view). If the item’s age is A before it is subjected to a PM action, then it has “virtual” age $A - x$ after the PM action where the reduction in the age is x , $0 \leq x \leq A$. The intensity function decreases after the PM action as shown in Figure 18.2.

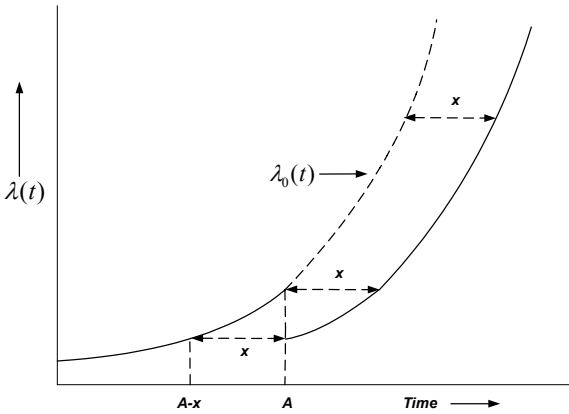


Figure 18.2. Effect of ‘age reduction’ PM actions

18.2.3.2 Two-dimensional Model Formulations

When item failure depends on age and usage, a 2-D failure model is needed. Two different approaches (one-dimensional and two-dimensional) to the failure modelling have been proposed in the literature.

One-dimensional approach

In the one-dimensional approach, the two-dimensional problem is effectively reduced to a one-dimensional problem by treating usage as a random function of age. A typical model assumes that $X(t)$, the usage of the item at age t , is a linear function of t , so

$$X(t) = \Gamma t \quad (18.6)$$

where the usage rate Γ , $0 \leq \Gamma < \infty$, is a non-negative random variable with distribution function $G(r)$ and density function $g(r)$.

The hazard function for time to time to first failure, conditional on $\Gamma = r$ is given by $h(t|r)$. Various forms of $h(t|r)$ have been proposed and one example is the polynomial function

$$h(t|r) = \theta_0 + \theta_1 r + \theta_2 t + \theta_3 X(t) + \theta_4 tX(t). \quad (18.7)$$

Most of the literature dealing with the one-dimensional approach assumes a linear relationship between usage and age; see for example, Blischke and Murthy (1994), Lawless *et al.* (1995), and Gertsbakh and Kordonsky (1998).

Two-dimensional approach

If T and X denote the item's age and usage at its first failure then, in the two-dimensional modelling approach, (T, X) is treated as a non-negative bivariate random variable with a bivariate distribution function. For more on this approach, see Murthy *et al.* (2006).

18.3 Warranties

18.3.1 Base Warranties

A base warranty (BW) is a contractual agreement between a customer (buyer) and a manufacturer (seller) that is entered into upon the sale of a product or service. It is bundled with the sale and its purpose is to establish the liability of the manufacturer in the event that an item fails or is unable to perform satisfactorily when properly used.

18.3.2 Classification of Base Warranties

A taxonomy for BW classification was first proposed by Blischke and Murthy (1994). The first classification criterion used is the requirement by the manufacturer to carry out further product development (for example, reliability improvement) subsequent to the product sale. Warranty policies that do not have this requirement can be further divided into two groups, the first consisting of those applicable to single item sales, and the second consisting of those used for the sale of groups of items (called lot or batch sales).

Policies in the first group can be subdivided into two sub-groups, depending on whether the policy is renewing or non-renewing. In a renewing policy, the warranty period begins anew with each failure, and in a non-renewing policy the replacement (or repaired) item assumes the residual warranty time of the item that failed. A further subdivision occurs by classifying warranties as “simple” or “combination”. Commonly used simple consumer BWs involve free replacement (FRW) and replacement at *pro-rata* cost (PRW). A combination policy combines the terms of two or more simple policies. Each of these four subgroups can be further subdivided based on whether the policy is 1-dimensional or two- (or higher) dimensional. A one-dimensional (1-D) policy is usually age-based but may sometimes be based on item usage. A two-dimensional (2-D) policy is based on time (age) as well as usage.

Repairable items are sold with either 1-D or 2-D non-renewing FRW policies. In the 1-D case, the manufacturer agrees to repair or provide replacements for failed items free of charge up to a time W from the time of the initial purchase and the BW expires at time W . A typical example is a one-year BW on a computer with the option for buyers to purchase extended warranty coverage. In the 2-D case, the manufacturer agrees to repair or provide a replacement for failed items free of charge up to a time W from the time of initial purchase (limit on time) or up to a usage U (limit on usage), whichever occurs first. A typical example is an automobile where the BW has a time limit of 2 years and a usage limit of 20,000 kilometres.

18.3.3 Warranty Servicing Cost Analysis

The two types of servicing costs of interest to the manufacturer are (1) expected cost per item sold, and (2) expected cost over the product life cycle.

18.3.3.1 Expected Cost per Item Sold

Whenever a failed item is returned for rectification action under warranty, the manufacturer incurs costs due to handling, material, labour, facilities, *etc.*, and these costs are random variables. The number of claims over the BW period is also random, so the total cost of servicing all the warranty claims is the sum of a random number of individual costs.

18.3.3.2 Expected Cost Over the Life Cycle of a Product

From the manufacturer's perspective, the product life cycle, L , is the period from the instant a new product is launched to the instant it is withdrawn from the market. During this period, product sales (first and repeat purchases) occur over time in a dynamic manner and the manufacturer must service all the warranty claims associated with each sale. For products sold with a 1-D non-renewing BW of length W , the total period for servicing claims is $L + W$. The total cost incurred by the manufacturer during this period depends on the servicing strategy used (the choice between repair and replacement) and the logistics of delivering the servicing.

From the customer's perspective, the life cycle cost (LCC) is the total cost of owning and operating a product over its useful life. This is of particular importance in the case of expensive products such as industrial plants, train systems, aircraft, *etc.*, that can be used for very long periods of time.

18.3.4 Extended Warranties

An extended warranty (EW) is a separate service contract that a customer may purchase to extend the warranty coverage when the BW expires. The EW price depends on the duration and the terms (not all parts might be covered, cost sharing, cost limits, *etc.*). EWs are offered not only by manufacturers but also by third parties such as retailers, insurance companies, *etc.* In most cases, the customer is required to purchase an EW at the time of purchase or just before the BW expires.

18.4 Link Between Warranty and Maintenance

In the case of BWs, the manufacturer incurs additional costs resulting from CM actions to rectify any warranty claims. For EWs, the costs are incurred by the EW provider. These additional costs result in increased prices for products and EWs and so are also of interest to customers.

If the useful life of a product is relatively short then so is its BW and warranty servicing should only involve CM actions. If a product has a long useful life, then an EW can also be relatively long and, in this case, the EW provider can reduce warranty servicing costs by performing effective PM. From the customer's perspective, PM and CM costs after both the BW and EW expire are of particular interest for certain products.

Hence, there is a close link between warranties (BW and EW) and maintenance (CM and PM) and there is an extensive literature that connects the two concepts. We first propose a taxonomy to classify this literature and then we provide a review.

18.4.1 Taxonomy for Classification

The literature involving warranty and maintenance can be organised into the following three categories:

- Warranty servicing with only CM actions. The papers in this category deal with the optimal choice of CM action (repair vs replace by new, different levels of repair) to minimise the expected warranty servicing cost per unit sale.
- Warranty servicing with both CM and PM actions. The papers in this category deal with the use of PM actions in order to achieve a proper trade-off between the additional PM costs and the reduction in the expected warranty servicing costs.
- Maintenance during the post-BW period. Here the focus is on the life cycle costs (LCC) and the role of the EW and PM actions during the post-BW period in order to minimise these costs.

18.4.2 Warranty Servicing Involving Only CM

Here the choice is between replace by new at each failure or repair and then the quality of repair to use in order to minimise the expected warranty cost. There are several papers dealing with this topic for 1-D warranties.

18.4.2.1 Minimal Repair

The first warranty servicing model with minimal repair (see Barlow and Hunter, 1960) was proposed by Nguyen (1984), who split the BW period into a replacement interval followed by a repair interval. The length of the first interval was chosen optimally in order to minimize the expected servicing cost.

Jack and Van der Duyn Schouten (2000) showed that Nguyen's (1984) servicing strategy was sub-optimal and conjectured that the optimal strategy is characterized in terms of three distinct intervals $[0, x]$, $[x, y]$, and $(y, W]$ where W is the length of the BW period. Minimal repairs are carried out in the first and third intervals and either minimal repair or replacement by new is used in the second interval depending on the item's age at failure. Because of the need to track the item's age, this strategy is difficult to implement and so Jack and Murthy (2001) proposed a close to optimal strategy with the same interval structure but with only the first item failure in the second interval resulting in a replacement and all other subsequent failures being minimally repaired. Jiang *et al.* (2006) proved the conjecture made by Jack and Van der Duyn Schouten (2000) to be true.

18.4.2.2 Different from New Repair

Biedenweg (1981) and Nguyen and Murthy (1986, 1989) also discussed strategies where the BW period is divided into distinct intervals for repair and replacement but assumed that repaired items have independent and identically distributed lifetimes different from that of a new item.

18.4.2.3 Imperfect Repair

Servicing strategies involving replacement of failed items are not appropriate when replacement costs are high compared to the cost of a minimal repair. In this case, it is more appropriate to use imperfect repair strategies (where the failure characteristics of the repaired item are better than those after minimal repair but are

not the same as a new item). The degree of reliability improvement after repairs are completed can be controlled by the manufacturer and so is a decision variable. The imperfect repair literature has focused mainly on its use in the context of PM and CM actions for unreliable systems. See, for example, Wang (2002), Li and Shaked (2003), Doyen and Gaudoin (2004), Zequeira and Berenguer (2006), Sheu *et al.* (2006), and Tang and Lam (2006). Chukova *et al.* (2004) look at warranty analysis with imperfect repairs.

Yun *et al.* (2008) consider two servicing strategies involving minimal and imperfect repairs. In both strategies a failed item is subjected to at most one imperfect repair over the BW period. In the first strategy, the level of reliability improvement under imperfect repair depends on the item age when the repair is carried out whereas, in the second strategy, the level of improvement is independent of the age. Consequently, the first strategy involves a functional optimisation to determine the optimal reliability improvement under an imperfect repair but only a parameter optimisation is involved in Strategy 2.

18.4.2.4 Two-dimensional Warranty Servicing

Servicing strategies for products sold with two-dimensional BWs have also been studied. Iskandar and Murthy (2003) discussed two strategies similar to those in Nguyen and Murthy (1986, 1989) but with minimal repair on failure. Iskandar *et al.* (2005) examined a servicing strategy similar to that given in Jack and Murthy (2001).

18.4.3 Warranty Servicing Involving Both CM and PM

CM must be performed throughout an item's useful life and PM may also be scheduled during and after the BW period. The following is a review of the warranty servicing literature where PM is also scheduled.

Chun and Lee (1992) considered the effect of performing periodic imperfect PM actions on an item during the BW period and the post-BW period. Each PM action reduced the item's age by a fixed amount and all failures between PM actions were minimally repaired. In the BW period, the manufacturer pays all the repair costs and a proportion of the cost of each PM action with the proportion depending on when the action is carried out. In the post-BW period, the customer pays for the cost of all repairs and PM actions. The optimal period between PM actions is obtained by minimising the customer's asymptotic expected cost per unit time over an infinite horizon.

Chun (1992) dealt with a similar problem to Chun and Lee (1992) but focused instead on the manufacturer's periodic PM strategy over the BW period. The optimal number of PM actions is obtained by minimising the expected cost of repairs and PM actions over this finite horizon.

Jack and Dagpunar (1994) showed that Chun's (1992) strictly periodic PM policy over the BW period is not optimal. They show that the optimal strategy is to perform a fixed number of periodic PM actions that, in each case, renew the item followed by a final interval of different length at the end of the BW period where only minimal repairs are carried out.

Dagpunar and Jack (1994) extended their previous model by assuming that the amount of age reduction is under the control of the manufacturer and the cost of each PM action depends on the item's age and on the effective age reduction resulting from the action. In their revised model, the optimal strategy can result in the item not being restored to as good as new at each PM action. The optimal number of PM actions, the optimal operating age at which to perform a PM action, and the optimal age reduction, are obtained by minimising the manufacturer's expected warranty servicing cost.

Sahin and Polatoglu (1996a) discussed two types of policy for item replacement during the post-BW period with all failures during the BW period being rectified at no cost to the customer. In the first policy, the item is replaced by a new item at a specified time after the BW ends. Failures before this time are minimally repaired with the customer paying for each repair. In the second policy, the replacement is postponed until the first failure after the specified time. Both stationary and non-stationary strategies are considered in order to minimise the customer's long run average cost. The non-stationary strategies depend on the information available to the customer at the end of the BW period regarding item age and number of previous failures. Sahin and Polatoglu (1996b) examined PM policies with uncertainty in product quality.

Monga and Zuo (1998) considered a model that included decisions about system, burn-in, warranty, and maintenance. They used genetic algorithms to determine the optimal values for system design, burn-in period, PM intervals, and replacement time by minimising the expected system life cycle cost. In their model, the manufacturer pays the costs of rectifying failures during the BW period and the customer pays post-BW costs.

Jung *et al.* (2000) found the optimal number and period for PM actions following the expiry of the BW by minimising the customer's asymptotic expected cost per unit time. Both the renewing PRW and the renewing FRW were considered. Jung and Park (2003) considered a similar model to Jung *et al.* (2000).

Jack and Murthy (2002) applied the intensity reduction method for imperfect PM modelling to determine the optimal number of PM actions and corresponding intensity reductions that a manufacturer should carry out to minimise expected servicing costs over the BW period.

Djamaludin *et al.* (2004) and Kim *et al.* (2001) introduced a framework to study preventive maintenance over a product's life cycle and proposed new models involving both continuous and discrete imperfect PM actions over fixed life cycles.

Pascual and Ortega (2006) extended the work of Kim *et al.* (2001) by proposing a model where optimal decisions on life-cycle duration and the number of imperfect PM actions to perform during it are made. The customer is also able to negotiate a longer BW period with the manufacturer by agreeing to perform PM during this interval.

18.5 Maintenance Logistics for Warranty Servicing

Maintenance logistics involves the planning by a manufacturer of all the relevant operations needed to service warranty claims. A manufacturer's ability to perform

warranty servicing is influenced by the geographical distribution of the customers and by their level of demand for prompt response to warranty claims. The manufacturer requires a dispersed network of service facilities to store spare parts and provide bases for field service. The service delivery network requires a diverse collection of human and capital resources and careful attention must be paid to both its design and control. Several strategic, tactical and operational issues are involved.

18.5.1 Strategic Issues

The main strategic issues are the location of warehouses and service centres, and the warranty servicing channels. The location of the service centres (to carry out repairs) and the warehouses (to stock necessary spares) depends on the geographical distribution of the customers who have purchased the product, the type of product, and its reliability characteristics.

18.5.1.1 Location of Service Centres

Most products have a complex structure and, when an item fails, the first task is to determine and identify the most likely cause of failure. For certain products (such as large home appliances or elevators in multi-story buildings), on-site diagnosis and repair are required. For others, the failed items are brought either to the retailer (for most consumer durables) or to some designated service centre. In the majority of cases, the failed item is made operational through appropriate actions at this level. However, in some instances, all failures at this level cannot be rectified due to the lack of resources such as specialised equipment and/or appropriately trained employees. The failed component must then be removed and transported to a higher-level service centre so that the rectification can be carried out.

There are often more than two levels, depending on the complexity of the product and the type of resources required to perform the rectification. For a jet engine, this might involve a service facility at a major airport (level 1) followed by a national (or regional) service centre (level 2) and finally a service centre at the plant where the engine was manufactured (level 3). If the item is repairable, the objective is to determine where the repair should take place in a multi-echelon repair facility.

Models to determine the number of service levels, the location of the service centres, and their capacities must take into account the following:

- The transportation time and cost to move failed and repaired items between service centres;
- The cost to operate the service centres (allowing for the equipment and skilled employees needed); and
- The capacity of each service centre which depends on the demand at the centre and this is determined by the geographical distribution of sales and the product reliability.

To solve the service centre location problem, the following topics need to be considered:

- Customer coverage (to ensure that all the customers can be reached);
- Distance that a failed item must travel to a service centre;
- Distance that a repairman has to travel for a field visit; and
- Given the coverage, models can be used to determine the demand at the centres and their capacities.

18.5.1.2 Location of Warehouses

Depending on the geographical area of the customers, a manufacturer might need to use a network of warehouses with a multi-echelon structure involving one or more levels. A multi-national manufacturer might have a regional warehouse (level 4) that receives parts from the different component manufacturers and then feeds these to national warehouses (level 3) for onward shipment to locally distributed warehouses (level 2) and then to service centres (level 1).

The problem of warehouse locations and their capacities has been discussed in the logistics literature but the existing models need to be modified to take into account the service centre locations and product reliability characteristics. The optimal locations must take the following into account:

- The transportation time and cost to move parts between warehouses (in the case of multi-echelon warehouses) and from warehouses to service centres;
- The operating cost of the warehouses; and
- The capacity of each warehouse based on the demand for spares at the various service centres supplied by the warehouse.

18.5.1.3 Service Channels

A manufacturer can select between the following two options for warranty servicing:

- The service is provided by retail or service centres owned and operated by the manufacturer; and
- The service is provided by an independent agent.

18.5.2 Tactical and Operational Issues

The tactical and operational issues in warranty logistics involve activities at the service centre level and decisions about spare part inventory levels, transportation of spares from warehouses to service centres, job scheduling, and repair vs replace decisions.

18.5.2.1 Spare Parts Inventory

The key decisions involved in spare parts inventories are the following:

- Which components should be carried as spare parts;
- The inventory levels of these parts; and
- The ordering frequencies for the parts and the amounts of parts that should be ordered.

These decisions depend on the expected numbers of component failures over time and these are influenced by sales levels and component reliability.

Most models dealing with spare part inventories have very simple assumptions about how inventory is depleted. In warranty-servicing, the depletion rate for a particular component is random and is influenced by the product sales rate across the region serviced by the servicing centre and product reliability. Optimal decisions about inventory levels and ordering policies need to take into these factors into account.

18.5.2.2 Material Transportation

Warranty servicing logistics involves material transportation (parts, failed items, etc.) from one location to another. The disciplines of materials management and of operations management discuss transportation problems (Tersine, 1994 and Nahmias, 1997) and many issues relating to transportation have been studied. Examples include integrating inventory and transportation (Qu *et al.* 1999) and emergency transshipments (Evers, 2001).

Three types of material transportation that are more specifically related to warranty servicing logistics are as follows:

- Transportation of failed units from a lower level to a higher level in a multi-echelon service structure;
- Transportation of repaired items from service centres to customers or pick-up points where they can be collected by customers; and
- Transportation of spares to and from warehouses.

The quantities to be shipped are random variables that are influenced by sales and product reliability and whether the shipping can be carried out either by the manufacturer or by an independent agent. In the latter case, a contract between the manufacturer and the independent agent must take into account transportation cost, transportation frequency, upper limits on transportation amounts, time limits, penalties for delivery delays and breaches of contract, etc.

Agency relationships to be discussed in Section 18.6 are also relevant in this case and different contract options may be evaluated using the Agency Theory framework.

18.5.2.3 Scheduling of Jobs, Repairs, and the Travelling Repairman Problem

Product support involves arrangements for repairing failed items. Products can be differentiated depending on whether failed items are brought to a service centre or a repairman needs to travel to rectify the failures. In the former case, job scheduling is an important issue that has an impact on the overall cost of providing the service and also on customer satisfaction. Job scheduling has been extensively studied (Hajri *et al.* 2000; Jianer and Miranda 2001; Ponnambalam *et al.* 2001). In the latter case, the problem is how to schedule the repair jobs to reduce travelling time and this is termed the “travelling repairman problem.” A number of solutions to the problem have been proposed (Afrati *et al.* 1986; Agnihothri 1998; Yang 1989). If a warranty includes penalties for service delays, then job scheduling also needs to take this into account.

18.6 Outsourcing of Maintenance for Warranty Servicing

A manufacturer or an EW provider can employ an independent agent to perform warranty servicing. The framework that is necessary to consider all the different issues involved with this outsourcing is provided by agency theory.

18.6.1 Agency Theory

Agency theory is concerned with the relationship that exists between two parties when one party (the principal) delegates work to be performed by a second party (the agent). A contract defines the relationship and agency theory helps to resolve the two problems that can take place.

The first problem occurs when the principal and the agent have conflicting goals and the principal finds it difficult or expensive to verify the agent's actions and whether or not the agent has behaved in a proper manner. The second problem occurs when the principal and the agent have different attitudes to risk and risk sharing takes place (due to various uncertainties).

The focus of agency theory, according to Eisenhardt (1989), is to determine the terms of the optimal contract, behaviour vs outcome, between the two parties. In the principal-agent literature, many different cases have been studied in depth and Figure 18.3 indicates the range of issues that have been covered. For an overview of the many different disciplines in which agency theory has been applied, see Acekere (1993).

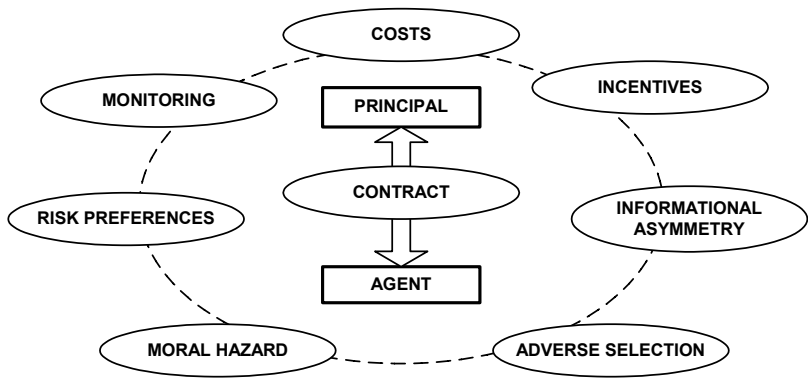


Figure 18.3. Agency theory issues

18.6.1.1 Issues in Agency Theory

Moral hazard:

Moral hazard refers to the agent's lack of effort in carrying out the delegated tasks. The two parties in the relationship have different objectives and the principal cannot assess the effort level that the agent has actually used.

Adverse selection:

Adverse selection refers to the agent misrepresenting their skills to carry out the tasks and the principal being unable to completely verify this before deciding to hire them.

Information:

To avoid adverse selection, the principal can try to obtain information about the agent's ability. One way of doing this is to contact people for whom the agent has previously provided service.

Monitoring:

The principal can counteract the moral hazard problem by closely monitoring the agent's actions.

Information asymmetry:

The overall outcome of the relationship is affected by several uncertainties. In general, the two parties will have different information to make an assessment of these uncertainties.

Risk:

This results from the different uncertainties that affect the outcome of the relationship. For a variety of reasons, the risk attitude of the two parties will differ and a problem arises when they disagree over the allocation of the risk.

Costs:

Both parties have various kinds of costs. Some of these depend on the outcome of the relationship (which is influenced by uncertainties), on acquiring information, monitoring, and on the administration of the contract. The centre of principal-agent theory lies the trade-off between (1) the cost of monitoring the agent's actions and (2) the cost of measuring the outcomes of the relationship and of transferring the risk to the agent.

Contract:

The design of the contract to take into account the above issues is the challenge that lies at the centre of the relationship between the principal and the agent.

For standard commercial and industrial products and also consumer durables, the terms of the EW policy are decided by the EW provider and the customer does not have any direct input. Agency Theory issues (such as moral hazard, adverse selection, risk, monitoring, *etc.*) are all relevant in the EW context. Current EWs offered lack flexibility for customers and many of these customers and also EW regulators believe that EW prices are too high. EW providers need to offer a menu of flexible warranties to meet the different needs of the customer population. Agency theory provides a framework to evaluate the costs of different EW policies taking into account all the relevant issues.

18.7 Conclusions and Topics for Future Research

Maintenance is an important concept in the context of warranties. This chapter has highlighted the link between the two subjects and the important issues involved have been discussed. Extensive literature reviews have also been provided.

There is scope for future research to be carried out in the following areas:

- Warranty servicing models which include both PM and CM for items covered by 2-D warranties;
- The use of agency theory to study the problems involved when warranty servicing is outsourced; and
- Specific analysis of spare parts inventories in the area of warranty logistics.

References

- Ackere A van (1993) The principal/agent paradigm: Its relevance to various fields. *Eur J of Oper Res* 70: 88–103.
- Afrati F, Cosmadakis S, Papadimitriou, CH, Papageorgiou G, and Papakostantinou N (1986) The complexity of the travelling repairman problem. *Inform Theor Appl* 20: 79–87
- Agnihothri SR (1998) Mean value analysis of the traveling repairman problem. *IIE Trans* 20: 223–229.
- Ascher H and Feingold H (1984) *Repairable System Reliability*, Marcel Dekker, New York
- Barlow RE and Hunter L (1960) Optimal preventive maintenance policies. *Oper Res* 8: 90–100.
- Biedenweg FM (1981) *Warranty Analysis: Consumer Value vs Manufacturers Cost*, Unpublished Ph.D. Thesis, Stanford University, U.S.A.
- Blischke WR and Murthy DNP (1994) *Warranty Cost Analysis*, Marcel Dekker, New York
- Blischke WR and Murthy DNP (2000), *Reliability*, Wiley, New York
- Cho ID and Parlar M (1991) A survey on maintenance models for multi-unit systems. *Eur J of Oper Res* 51: 1–23.
- Chukova S, Arnold R and Wang DQ (2004), *Warranty analysis: An approach to modeling imperfect repairs*. *Int J of Prod Econ* 89: 57–68
- Chun YH (1992) Optimal number of periodic preventive maintenance operations under warranty. *Reliab Eng and Syst Saf* 37: 223–225.
- Chun YH and Lee CS (1992) Optimal replacement policy for a warrantied system with imperfect preventive maintenance operations. *Microelectroni and Reliab* 32: 839–843.
- Dagpunar JS and Jack N (1994) Preventive maintenance strategy for equipment under warranty. *Microelectroni and Reliabi* 34: 1089–1093.
- Dekker R and Scarf PA (1998) On the impact of optimisation models in maintenance decision making: the state of the art. *Reliab Eng and Syst Saf* 60: 111–119.
- Djamaludin I, Murthy DNP and Kim CS (2004) Warranty and discrete preventive maintenance. *Reliab Eng and Syst Saf* 84: 163–168.
- Doyen L and Gaudoin O (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliab Eng and Syst Saf* 84: 45–56
- Eisenhardt KM (1989) Agency theory: An assessment and review. *Acad of Manag Rev* 14: 57–74.
- Evers PT (2001) Heuristics for assessing emergency transshipments. *Eur J of Oper Res* 129: 311–316

- Gertsbakh IB and Kordonsky KB (1998) Parallel time scales and two-dimensional manufacturer and individual customer warranties. *IEE Trans* 30: 1181–1189.
- Hajri S, Liouane N, Hammadi S, and Borne P (2000) A controlled genetic algorithm by fuzzy logic and belief functions for job-shop scheduling. *IEEE Trans on Syst Man and Cybernet Part B (Cybernet)* 30: 812–818.
- Iskandar BP, and Murthy DNP (2003) Repair-replace strategies for two-dimensional warranty policies. *Math and Comput Model* 38: 1233–1241.
- Iskandar BP, Murthy DNP and Jack N (2005) A new repair-replace strategy for items sold with a two-dimensional warranty. *Comput and Oper Res* 32: 669–682.
- Jack N and Dagpunar JS (1994) An optimal imperfect maintenance policy over a warranty period. *Microelectroni and Reliab* 34: 529–534
- Jack N. and Murthy DNP (2001) A servicing strategy for items sold under warranty. *J of the Oper Res Soc* 52: 1284–1288.
- Jack N. and Murthy DNP (2002) A new preventive maintenance strategy for items sold under warranty. *IMA J of Manag Math* 13: 121–129.
- Jack N and Van der Duyn Schouten F (2000) Optimal repair-replace strategies for a warranted product. *Int J of Prod Econ* 67: 95–100
- Jianer C and Miranda A (2001) A polynomial time approximation scheme for general multiprocessor job scheduling. *SIAM J on Comput* 31: 1–17
- Jiang X, Jardine AKS, and Lugitigheid D (2006) On a conjecture of optimal repair-replacement strategies for warranted products. *Math and Comput Model* 44:963–972.
- Jung GM and Park DH (2003) Optimal maintenance policies during the post-warranty period. *Reliab Eng and Syst Saf* 82: 173–185.
- Jung GM, Lee CH and Park DH (2000) Periodic preventive maintenance policies following the expiration of warranty. *Asia-Pac J of Oper Res* 17: 17–26
- Kim CS, Djameludin I and Murthy DNP (2001) Warranty and preventive maintenance. *Int J of Reliab Qual and Saf Eng* 8: 89–107.
- Lawless J, Hu J and Cao J (1995) Methods for estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis* 1: 227–240
- Li H and Shaked M (2003) Imperfect repair models with preventive maintenance. *J of Appl Probab* 40: 1043–1059
- Meeker WQ and Escobar LA (1998) *Statistical Methods For Reliability Data*, Wiley, New York
- Monga A and Zuo MJ (1998) Optimal system design considering maintenance and warranty. *Comput Oper Res* 9: 691–705
- Murthy DNP, Baik J, Wilson RJ and Bulmer M (2006) Two-dimensional Failure Modelling, in *Springer Handbook of Engineering Statistics*, H. Pham [Editor], Springer-Verlag, London, pp 97–111
- Nahmias S (1997) *Production and Operational Analysis*, McGraw-Hill, New York.
- Nguyen DG (1984) *Studies in warranty policies and product reliability*. Unpublished Ph.D. Thesis, The University of Queensland, Australia.
- Nguyen DG and Murthy DNP (1986) An optimal policy for servicing warranty. *J Oper Res Soc* 37: 1081–1088.
- Nguyen DG and Murthy DNP (1989) Optimal replace-repair strategy for servicing items sold with warranty. *Eur J of Oper Res* 39: 206–212.
- Pascual R and Ortega JH (2006) Optimal replacement and overhaul decisions with imperfect maintenance and warranty contracts. *Reliab Eng and Syst Saf* 91: 241–248.
- Ponnambalam SG, Ramkumar V and Jawahar N (2001) A multiobjective genetic algorithm for job shop scheduling. *Prod Planning and Control* 12: 764–774
- Qu W W, Bookbinder JH and Iyogun P (1999) An integrated inventory-transportation system with modified periodic policy for multiple products. *Eur J of Oper Res* 115: 254–269.

- Sahin I and Polatoglu H (1996a) Maintenance strategies following the expiration of warranty. *IEEE Trans on Reliab* 45: 220–228
- Sahin I, Polatoglu H (1996b) Manufacturing Quality, Reliability and preventive maintenance. *Prod and Oper Manag* 5: 132–147.
- Sheu SH, Lin YB and Liao GL (2006) Optimum policies for a system with general imperfect maintenance. *Reliab Eng and Syst Saf* 91, 362–369.
- Tang YY and Lam Y (2006) A δ -shock maintenance model for a deteriorating system. *Eur J of Oper Res* 168: 541–556
- Tersine RJ (1994) *Principles of Inventory and Materials Management*. Prentice Hall, Englewood Cliffs, NJ
- Wang H (2002) A survey of maintenance policies of deteriorating systems. *Eur J of Oper Res* 139: 469–489
- Yang CE (1989) A dynamic programming algorithm for the traveling repairman problem. *Asia-Pac J of Oper Res* 6: 192–206
- Yano CA (1992) Optimizing transportation contracts to supply just-in-time deliveries: The case of one contracted vehicle per shipment. *IIE Trans* 24: 177–183.
- Yun W-Y, Murthy DNP and Jack N (2008) Warranty servicing with imperfect repair. *Int J of Prod Econ* 111: 159–169
- Zequeira RI and Bérenguer C (2006) Periodic imperfect preventive maintenance with two categories of competing failure modes. *Reliab Eng and Syst Saf* 91: 460–468.

Delay Time Modeling for Optimized Inspection Intervals of Production Plant

Wenbin Wang

19.1 Introduction

Periodic inspections remain as one of the effective maintenance strategies currently used in industry. One of the key decision variables in such a strategy is the determination of the inspection intervals that can be regular or irregular. To clarify the objective of the type of the maintenance strategy we are concerned with here, consider a plant item with a maintenance strategy of inspecting every period T hours, days, weeks, months, ... , with repair of failures undertaken as they arise. The inspection consists of a check list of activities to be undertaken, and a general inspection of the operational state of the plant. Any defect identified leads to immediate repair, and the objective of the maintenance strategy is to minimise operational downtime. Other objectives could be considered, cost, availability, output, ..., but for now we consider downtime reduction.

Conceptually, there is a relationship between the expected downtime per unit time $D(T)$, and the service period T ; see Figure 19.1. If T was small, the downtime per unit time would be large because the plant would frequently be unavailable due to servicing, and if T was sufficiently large, the downtime per unit time would essentially be that under a breakdown maintenance policy. If the chosen service period is T^* , all that can be expected to be known of $D(T)$ is the observed value $D(T^*)$, that is the current downtime measure. One wishes to reduce $D(T^*)$, and if a model such as Figure 19.1 were available, there would be little difficulty in identifying a good operational period for T , which could be infinite, that is, do not inspect. Unfortunately, in the absence of modeling, all that is generally available is the data of Figure 19.2.

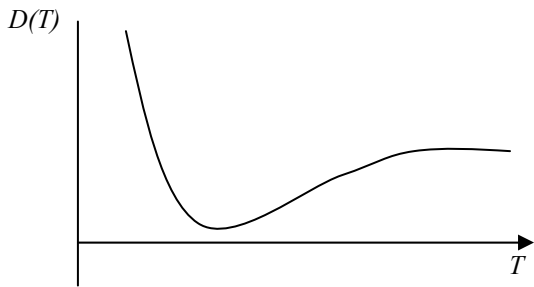


Figure 19.1. Downtime model: inspection period T

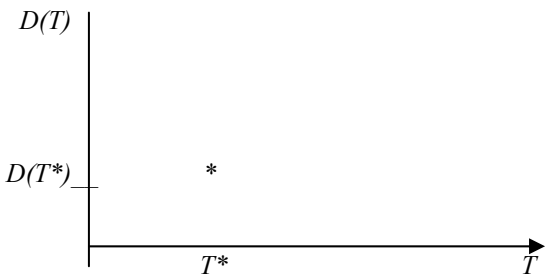


Figure 19.2. Downtime information

To move from Figure 19.2 to Figure 19.1 requires maintenance modeling, and Figure 19.1 is a graphical representation of the model. A modeling tool which can be used to build such a relationship presented in Figure 19.1 is the well known delay time (DT) modeling technique which was first mentioned in Christer (1976) and subsequently developed by Christer and Waller (1984), Baker and Wang (1991), Christer *et al.* (1995), Christer (1999), Wang and Christer (2003), Wang and Jia (2007), and Akbarov *et al.* (2008).

This chapter deals with the modeling, analysis and optimization of such an inspection problem using the DT concept. This concept provides a modeling framework readily applicable to a wide class of actual industrial maintenance problems. The chapter is organized as follows. Section 19.2 presents an introduction to the DT concept. Section 19.3 introduces the DT models for complex plant. Section 19.4 focuses on DT model parameters estimation. Section 19.5 presents a case example and Section 19.6 outlines some other developments and future research on DT modeling.

19.2 The DT Concept and Modeling Characteristics

We are interested in the relationship between the performance of equipment and maintenance intervention, and to capture this the conventional reliability analysis of time to first failure, or time between failures, requires enrichment. Consider a repairable item of plant. It could be, say, a component, a machine, or an integrated

set of machines forming a production line, but viewed by management as a plant unit. The interaction between maintenance concept and equipment performance may be captured using the DT concept presented below.

Let the item of plant be maintained on a breakdown basis. The time history of breakdown or failure events is a random series of points; see Figure 19.3. For any one of these failures, the likelihood is that had the plant been inspected at some point just prior to failure, it could have been seen that all was not well and a defect was present which, though the plant was still working, would ultimately lead to a failure. Such signals include excessive vibration, unusual noise, excessive heat, surface staining, smell, reduced output, increased quality variability, The first instance where the presence of a defect might reasonably be expected to be recognised by an inspection had it taken place is called the initial point u of the defect, and the time h to failure from u is called the delay time of the defect; see Figure 19.4. Had an inspection taken place in $(u, u+h)$, the presence of a defect could have been noted and corrective actions taken prior to failure. Given that a defect arises, its delay time represents a window of opportunity for preventing a failure. Clearly, the delay time h is a characteristic of the plant concerned, the type of defect, the nature of any inspection, and perhaps the person inspecting. For example, if the plant was a vehicle, and the maintenance practice was to respond when the driver reported a problem, then there is in effect a form of continuous monitoring inspection of cab related aspects of the vehicle, with a reasonably long delay time consistent with the rate of deterioration of the defect. However, should the exhaust collapse because a support bracket was corroded through, the likely warning period for the driver, the delay time, would be virtually zero, since he would not normally be expected to look under the vehicle. At the same time, had an inspection been undertaken by a service mechanic, the delay time may have been measured in weeks or months. Had the exhaust collapsed because securing bolts became loose before falling out, then the driver could have had a warning period of excessive vibration, and perhaps noise, and the defects of a drive related delay time measured in days or weeks.

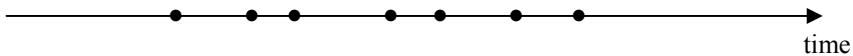


Figure 19.3. Failure points ‘•’

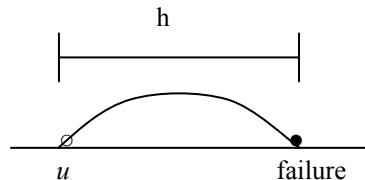


Figure 19.4. The delay time for a defect

To see why the delay time concept is of use, consider Figure 19.5 incorporating the same failure point pattern as Figure 19.3 along with the initial points associated with each failure arising under a breakdown system. Had an inspection taken place at point A, one defect could have been identified and the seven failures could have

been reduced to six. Likewise, had inspection taken place at point B and point A, four defects could have been identified and the seven failures could have been reduced to three. Figure 19.5 demonstrated that provided it is possible to model the way defects arise, that is the rate of arrival of defects $\lambda(u)$, and their associated delay time h , then the delay time concept can capture the relationship between inspection frequency and the number of plant failures.

We are assuming for now that inspections are perfect, that is, a defect is recognised if it is there and only if it is there, and is removed by a corrective action. DT modeling is still possible if these assumptions are not valid, but this more complex case is discussed in a subsequent section.

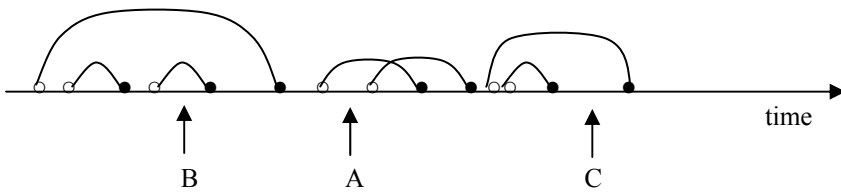


Figure 19.5. ‘o’ initial points; ‘•’ failure points

To put the above situation into the framework of modeling, the following are the characteristics of modeling a piece of production plant using the DT concept:

- Many failures can be characterized by a two-stage failure process, that is, from new to the initial point of a defect, and from this point to failure if the defect was not attended to;
- The initial points of defects are random and as such can be modeled by a stochastic process along the time axis;
- The time interval between the initial point of the defect and failure is uncertain and can be modeled by a probability distribution function;
- By inspections at discrete points, one can detect if the defect has appeared and then maintenance decisions can be initiated to avoid failure – this is preventive maintenance (PM) action;
- Failures can be observed immediately and need to be rectified through corrective maintenance (CM) actions at the time of the failure;
- All actions (inspection, PM and CM actions) may cost money and result in downtime; and
- The problem under study is to decide on the optimal inspection intervals that can be periodic or non-periodic.

One needs to build models to determine the optimal inspection intervals based on some objective function (or performance measure) – either downtime or cost or reliability.

19.3 The DT Models for Complex Plant

A complex plant, or multi-component plant, is one where a large number of failure modes arise, and the correction of one defect or failure has nominal impact in the steady state upon the overall plant failure characteristics. Consider the following complex plant maintenance modeling scenario where:

1. An inspection takes place every T time units, costs c_s units and requires d_s time units, where $d_s \ll T$;
2. Defects identified will be repaired during the inspection period;
3. Failure will be repaired immediately at an average cost c_f and downtime d_f ;
4. The plant has operated sufficiently long since new to be effectively in a steady state; and
5. Defects and failures only arise whilst plant is operating.

These assumptions characterise the simplest non-trivial inspection maintenance problem and would, of course, only be agreed in any particular case after careful analysis and investigation of the specific situation. We can now proceed to construct the mathematical model of the conceptual model in Figure 19.1.

19.3.1 The Down Time/Cost Model

Define $E[N_f((i-1)T, iT)]$ – expected number of failures over $[(i-1)T, iT]$, and $E[N_s(iT)]$ – expected number of defects identified at iT where $i=1,2,\dots$; then we have the downtime model given by

$$D(T) = \frac{d_f E[N_f((i-1)T, iT)] + d_s}{T + d_s} \quad (19.1)$$

The cost model follows immediately if we replace the downtime parameters by the cost parameters in Equation 19.1, that is,

$$C(T) = \frac{c_f E[N_f((i-1)T, iT)] + c_s}{T + d_s} \quad (19.2)$$

Equation 19.1 or 19.2 is established assuming that the defects identified at an inspection will always be removed without costing any extra downtime or cost. This assumption can be relaxed. Let d_r be the mean downtime per defect being repaired at an inspection. Then using the same approach as before, the expected downtime is given by

$$D(T) = \frac{d_f E[N_f((i-1)T, iT)] + d_s + d_r E[N_s(iT)]}{T + d_s + d_r E[N_s(iT)]} \quad (19.3)$$

Equation 19.1 or 19.3 is the algebraic form of Figure 19.1, where $E[N_f((i-1)T, iT)]$ and $E[N_s(iT)]$ will be given later. This measure gives the ratio of downtime (cost) to the total cycle time.

Now the key elements in Equations 19.1 and 19.3 are the derivations of $E[N_f((i-1)T, iT)]$ and $E[N_s(iT)]$, which will be presented in the next subsections depending on whether the inspection is perfect or not.

19.3.2 Modeling $E[N_f((i-1)T, iT)]$ and $E[N_s(iT)]$ Under the Assumption of Perfect Inspections

First we introduce some further assumptions and notation:

1. Inspections are perfect in that all (and only) defects present are identified;
2. Defects arise according to a Homogeneous Poisson Process (HPP) with the rate of occurrence of defects, λ , per unit time; and
3. The random delay time, H , of a random defect is described by a pdf $f(h)$, cdf $F(h)$, where h is the realisation of H , and is independent of the initial point U .

Since in this case both $E[N_f((i-1)T, iT)]$ and $E[N_s(iT)]$ are identical over each inspection interval because of the assumption of perfect inspections, we can simply denote $E[N_f(T)] = E[N_f((i-1)T, iT)]$ and $E[N_s(T)] = E[N_s(iT)]$.

It can be shown that the failure process shown in Figure 19.5 is a Marked Poisson process (Taylor and Karlin, 1998), with the delay time h as the marker. It has been proved that this failure process over $[0, T]$ is a nonhomogenous Poisson process (NHPP) (Taylor and Karlin, 1998; Christer and Wang, 1995). To derive the rate of occurrence of failures (ROCOF), $\nu(t)$, for this NHPP, within $[0, T]$, we start first by deriving the expected number of failures within $[0, T]$. Since the expected number of the defects arrived within $[t, t + \delta t]$, $0 \leq t < T$, is $\lambda \delta t$, then the expected value of the failures caused by these defects is $\lambda F(T-t) \delta t$. Integrating t from 0 to T and after some manipulation we have

$$E[N_f(T)] = \int_0^T \lambda F(t) dt \quad (19.4)$$

Differentiating (4) with respect to T we have

$$\nu(t) = \lambda F(t) \quad (19.5)$$

It can be shown that $N_s(T)$ also follows a Poisson distribution, Christer and Wang (1995), with the mean given by

$$E[N_f(T)] = \int_0^T \lambda(1 - F(t))dt \quad (19.6)$$

Summing Equations 19.4 and 19.5 we have the expected number of defects within $[0, T)$ is given, as expected, by,

$$E(N_d(T)) = \lambda T \quad (19.7)$$

Example 19.1: Suppose the pdf. of delay time was given by $f(h) = \alpha e^{-\alpha h}$. Then, from Equation 19.4 we have $E[N_f(T)] = T\lambda + \frac{\lambda}{\alpha}(e^{-\alpha T} - 1)$. Therefore the downtime per unit time model, Equation 19.1, becomes

$$D(T) = \frac{d_f(T\lambda + \frac{\lambda}{\alpha}(e^{-\alpha T} - 1)) + d_s}{T + d_s}$$

This may be re-arranged more usefully into

$$D(T) = \left\{ \frac{\left(d_s - \frac{\lambda d_f}{\alpha}\right) + \lambda T d_f + \frac{\lambda}{\alpha} e^{-\alpha T} d_f}{T + d_s} \right\}$$

in which the numerator depicts a constant term, a linear term in T , and a damped exponential term. We can now see that as $T \rightarrow \infty$, that is the cycle times become very large effectively eliminating any effect of inspection, then $D(T) \rightarrow \lambda d_f$ as expected. This simply states that as inspections get so far apart as effectively to not exist, every defect leads to a breakdown with downtime d_f , so the downtime per unit time is λd_f .

Again, if $T \rightarrow 0$, that is inspections are repeated very rapidly with little or no opportunity to operate the plant, then $D(T) \rightarrow 1$, that is the plant is always in the state of experiencing downtime d_s , so the downtime per unit time is maximum at unity.

19.3.3 Modeling $E[N_f((i-1)T, iT)]$ and $E[N_s(iT)]$ Under the Assumption of Imperfect Inspections

All the assumptions proposed in Section 19.3.1 will hold except the perfect inspection one. Assume for now that if a defect is present at an inspection, then there is a probability r that the defect can be identified. This implies that there is a probability $1-r$ that the defect will be unnoticed. Figure 19.6 depicts such a process.

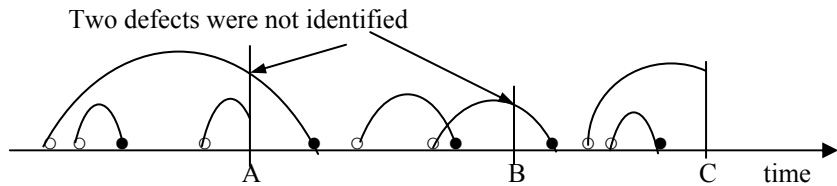


Figure 19.6. Failure process of a complex system subject to three non-perfect inspections at points A, B, and C, and two potential failures were removed and two missed

Define $v_i(t)$ --- ROCOF at time t , $t \in [(i-1)T, iT]$, it can be shown (Christer *et al.*, 1995; Christer and Wang, 1995), that $v_i(t)$ is given by

$$v_i(t) = \lambda \sum_{n=1}^i (1-r)^{i-n+1} [F(t-(n-1)T) - F(t-nT)] + \lambda F(t-(i-1)T) \tag{19.8}$$

for $t \in [(i-1)T, iT]$.

It can also be proved by induction that $v_{i-1}(t) \approx v_i(t)$ when i is large. Given that Equation 19.8 is available, it is straightforward that the expected number of failures over $[(i-1)T, iT]$ is given by

$$\begin{aligned} E[N_f((i-1)T, iT)] &= \int_{(i-1)T}^{iT} v_i(t) dt \\ &= \int_{(i-1)T}^{iT} \left\{ \lambda \sum_{n=1}^i (1-r)^{i-n+1} [F(t-(n-1)T) - F(t-nT)] + \lambda F(t-(i-1)T) \right\} dt \end{aligned} \tag{19.9}$$

The expected number of defects found at an inspection point, say, iT , is also a Poisson variable with the mean given by (Christer *et al.*, 1995; Christer and Wang, 1995)

$$E[N_s(iT)] = \lambda \sum_{n=1}^i (1-r)^{i-n+1} r \int_{(n-1)T}^{nT} [1 - F(iT-u)] du + \lambda r \int_{(i-1)T}^{iT} [1 - F(iT-u)] du \tag{19.10}$$

Example 19.2: Assuming the rate of occurrence of defects is two per day, and the delay time distribution is exponential with scale parameter 0.03 measured in days. The downtime measures are $d_f = 30$ and $d_p = 30$ min respectively. The probability of a perfect inspection is assumed to be 0.7. Using Equations 19.9, 19.10 and 19.3, we have the expected downtime against inspection intervals shown in Figure 19.7. It can be seen from Figure 19.7 that a weekly inspection interval is the best.

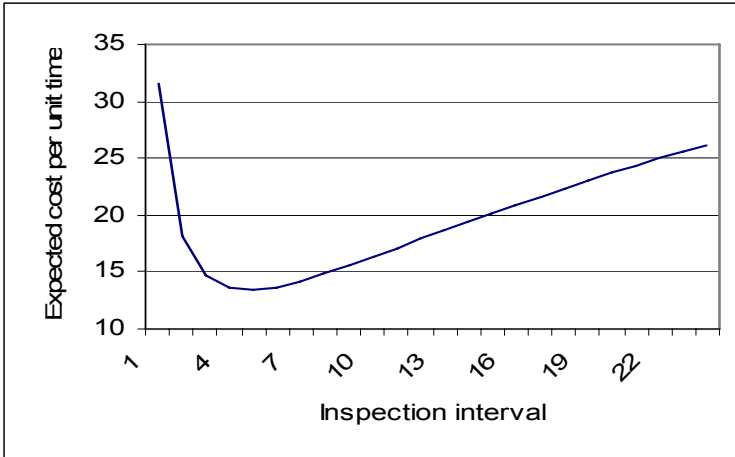


Figure 19.7. Expected downtime per unit time vs inspection interval (in days)

19.4 Delay Time Model Parameters Estimation

19.4.1 Introduction

In previous sections, delay time models for complex systems have been introduced. However, in a practical situation, before the construction of expected cost or downtime models, it is necessary to estimate the values of the parameters that characterise the defect arrival and failure processes. In this section we discuss the approaches that have been developed to estimate the parameters from ‘objective’ data collected at failures and inspections. In order to estimate the underlying parameters to any degree of accuracy, we require the availability of a sufficient number of cycles of data. In this section we discuss means by which different forms of delay time models incorporating different choices of delay time distribution can be compared and their fit to the objective data established.

Naturally, the parameter estimation process is not the same for the different types of delay-time model, *i.e.*, single component models where a single potential failure state is modeled and only one defect may (or may not) be present at any one time, compared with complex system models where many defects can exist simultaneously and many failures can occur in the interval between inspections. However, despite the modeling differences, the parameters are estimated on the same basis: using the failure/inspection data and the form of a specified delay time model, we develop the probability of observing each piece of information or event. In some cases, the probability of one event is conditional on previous events having already occurred.

Using maximum likelihood estimation (MLE), we are able to develop an expression incorporating the probabilities associated with all observed events in the data. The resulting likelihood function is then optimized with respect to the

parameters to obtain the estimated values; this involves finding those estimates that when inserted into the likelihood function give the ‘maximum likelihood’. This process can be simplified by taking natural logarithms of the likelihood function as maximising the log-likelihood will produce the same parameter estimates as the standard likelihood function. We introduce in this chapter only the parameter estimation techniques for complex systems.

19.4.2 Complex System – Parameter Estimation

Objective data for complex systems under regular inspections should consist of the failures (and associated times) in each interval of operation between inspections and the number of defects found in the system at each inspection. A typical scenario is illustrated in Figure 19.8 for a process with non-perfect inspections at time $iT, (i+1)T, (i+2)T$.

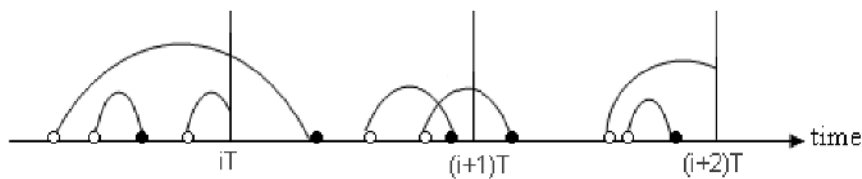


Figure 19.8. Illustrating a typical underlying defect-to-failure and non-perfect inspection process

However, all the information that we have to build a model of the system is illustrated in Figure 19.9.

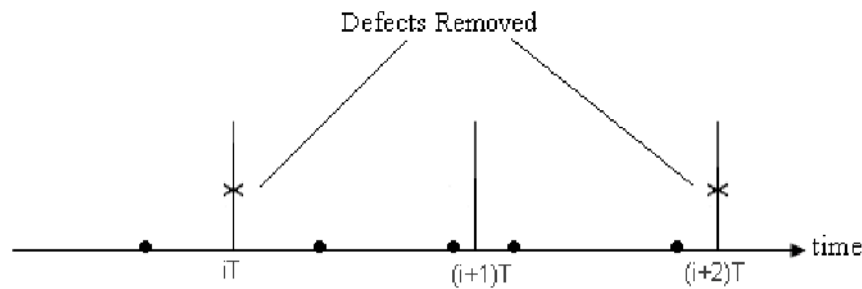


Figure 19.9. Illustrating the actual observed data for the process depicted in Figure 19.8

From this information, we estimate the parameters for the chosen form of the delay time model.

For all the different delay-time models discussed in previous sections, the number of failures arising over a constant interval between inspections, $[(i-1)T, iT)$, adheres to a non-homogenous Poisson process (NHPP) with $\nu_i(t)$ and the means given by Equations 19.9 and 19.10. We now consider the cases for perfect and impact inspection respectively.

19.4.2.1 Estimation Under the Perfect Inspection Assumption

Initially we consider the simple case of the estimation problem for the perfect inspection DT model with exponentially distributed delay times where only the number of failures, m_i , occurring in each cycle $[(i-1)T, iT)$ and the number of defects found and repaired, j_i , at each inspection iT are required. We do not need the actual failure times within the cycles for estimation of this perfect inspection DT model parameters.

The probability of observing m_i failures in $[(i-1)T, iT)$ is

$$P(N_f((i-1)T, iT) = m_i) = \frac{e^{-E[N_f(T)]} E[N_f(T)]^{m_i}}{m_i!} \quad (19.11)$$

Similarly the probability of removing j_i defects at inspection iT is

$$P(N_s(iT) = j_i) = \frac{e^{-E[N_s(T)]} E[N_s(T)]^{j_i}}{j_i!} \quad (19.12)$$

As the observations are independent, the likelihood of observing the given data set is just the product of the Poisson probability of observing each cycle of data, m_i and j_i . As such, the likelihood function for K inspection intervals of data is

$$\begin{aligned} \text{Likelihood} &= \prod_{i=1}^K \{P(N_f((i-1)T, iT) = m_i) P(N_p(iT) = j_i)\} \\ &= \prod_{i=1}^K \left\{ \left(\frac{e^{-E[N_f(T)]} E[N_f(T)]^{m_i}}{m_i!} \right) \left(\frac{e^{-E[N_p(T)]} E[N_p(T)]^{j_i}}{j_i!} \right) \right\} \end{aligned} \quad (19.13)$$

The likelihood function is optimized with respect to the parameters to obtain the estimated values. This process can be simplified by taking natural logarithms. The log-likelihood function is;

$$\begin{aligned} \ell &= \left(\sum_{i=1}^K m_i \right) \log(E[N_f(T)]) + \left(\sum_{i=1}^K j_i \right) \log(E[N_s(T)]) - K(E[N_f(T)] + E[N_s(T)]) \dots \\ &\quad \dots + \sum_{i=1}^K (\log(m_i!) + \log(j_i!)) \end{aligned} \quad (19.14)$$

where the final summation term is irrelevant when maximizing the log-likelihood as it is a constant term and therefore not a function of any of the parameters under investigation.

Consider the perfect inspection delay time model with a constant rate of defect arrival, λ , and exponentially distributed delay times with scale parameter α . From Section 19.3, Example 19.1, we know that $E[N_f(T)] = T\lambda + \frac{\lambda}{\alpha}(e^{-\alpha T} - 1)$ and inserting the expressions for $E[N_f(T)]$ and $E[N_s(T)]$ into the log-likelihood function, Equation 19.14, we obtain the following expression for the basic delay time model:

$$\ell = \left(\sum_{i=1}^K m_i \right) \log \left(\lambda T - \frac{\lambda}{\alpha} (1 - e^{-\alpha T}) \right) + \left(\sum_{i=1}^K j_i \right) \log \left(\frac{\lambda}{\alpha} (1 - e^{-\alpha T}) \right) - K \lambda T \quad (19.15)$$

As can be seen the log-likelihood for the model demands only that we know the total number of failures and defects removed and not necessarily their respective times. In a practical scenario, one would also know the number of cycles of data K and the length of the interval between inspections T .

For most of the complex system delay time models that are covered in previous sections, parameter estimation requires the use of an optimization algorithm to find the estimates that give the maximum value of the log-likelihood function. This is because the modeling for these cases requires the estimation of three or more parameters to represent the failure process and an analytical solution is not available. However, the delay time model introduced above contains only two unknown parameters in the rate of arrival λ and the scale parameter α and can be solved by partial differentiation with respect to λ and a simple line search for the remaining parameter α .

Example 19.3: with the following parameters: $\lambda = 0.05$ per hour and $\alpha = 0.05$ we simulate the failure and inspection process with an interval between inspections of $T = 60$ h for $K = 50$ inspection cycles. In the resulting output we observe 102 failures and 48 defects removed at inspections. Using the output, we attempt to recapture the parameters λ and α .

From Equation 19.15 we have the log-likelihood of observing the data

$$\ell = (102) \log \left(60\lambda - \frac{\lambda}{\alpha} (1 - e^{-60\alpha}) \right) + (48) \log \left(\frac{\lambda}{\alpha} (1 - e^{-60\alpha}) \right) - (60 \times 50) \lambda$$

Taking the partial differential of the log-likelihood with respect to λ and equating with 0 we obtain

$$\frac{\partial \ell}{\partial \lambda} = \frac{102 \left(60 - \frac{1}{\alpha} (1 - e^{-60\alpha}) \right)}{\lambda \left(60 - \frac{1}{\alpha} (1 - e^{-60\alpha}) \right)} + \frac{48 \left(\frac{1}{\alpha} (1 - e^{-60\alpha}) \right)}{\lambda \left(\frac{1}{\alpha} (1 - e^{-60\alpha}) \right)} - (60 \times 50) = 0$$

By cancellation of the terms in brackets for the numerator and denominator of each fraction

$$\hat{\lambda} = \frac{102 + 48}{60 \times 50} = 0.05$$

We could have arrived at the result for $\hat{\lambda}$ through a common-sense approach. Each event, whether a failure or defect removed at inspection, represents the outcome of one defect. As we are only interested in the average rate of arrival, it is obvious that this is given by

$$\hat{\lambda} = \left(\frac{\text{Total failures and defect removals over all cycles}}{\text{Total time over all cycles}} \right)$$

Having estimated the value of λ , we have to search for an estimate of α by using Equation 19.15 since the estimate of λ is known. This can be done easily

using any optimisation routine or even by an exhaustive search. For example substitute $\lambda=0.05$ into Equation 19.15 and plot the value of the likelihood; in Figure 19.10 we can that $\alpha=0.05$ maximises the likelihood, which is actually the true value.

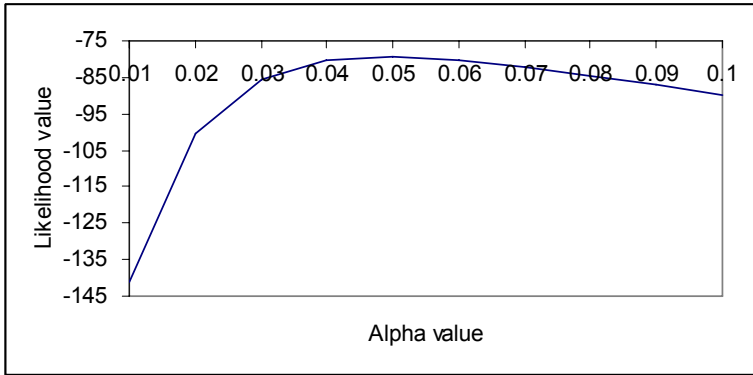


Figure 19.10. Plot of the likelihood of Equation 19.15 in terms of α

19.4.2.2 Estimation Under the Imperfect Inspection Assumption

As discussed in Section 19.3, the assumption of the perfect DT model is often not justified in practical scenarios and more advanced delay time models are needed to represent the failure and inspection process. When estimating the parameters for imperfect DT models, it is often necessary to refine the likelihood function (Equation 19.14) by considering the detailed pattern of behaviour within each interval in terms of the number of failures in smaller increments of the intervals. In the non-perfect inspection case a greater number of failures would occur earlier in the interval as some defects from previous cycles would have avoided detection and remained in the system.

To refine the likelihood function, the interval $[0, T)$ is broken down into z non-overlapping increments of duration θ :

$$I_b^i = [((i-1)T + (b-1)\theta), ((i-1)T + b\theta)] \quad (19.16)$$

If iT is the time of the i th inspection then

$$(i-1)T + z\theta = iT$$

The number of failures in each increment b of every interval i is Poisson distributed with mean $E[N_f(I_b^i)]$ and the probability of observing m_{ib} failures in I_b^i is

$$P(N_f(I_b^i) = m_{ib}) = \frac{E[N_f(I_b^i)]^{m_{ib}} e^{-E[N_f(I_b^i)]}}{m_{ib}!} \quad (19.17)$$

Given that Equation 19.8 is available, then $E[N_f(I_b^i)]$ can be obtained by integrating Equation 19.8 over the interval of I_b^i . The probability of observing a specific number of failures in an increment is homogenous across all intervals in the sample if the system is in a steady state. Note that we require the failure time of each breakdown within the interval. This is often observed in practice where, for instance, failures may be recorded as X h/days/weeks, *etc.*, after the last inspection as shown in Figure 19.11, where the data could be grouped data prepared for steady state analysis or it could pertain to a single interval in the refined interval estimation case.

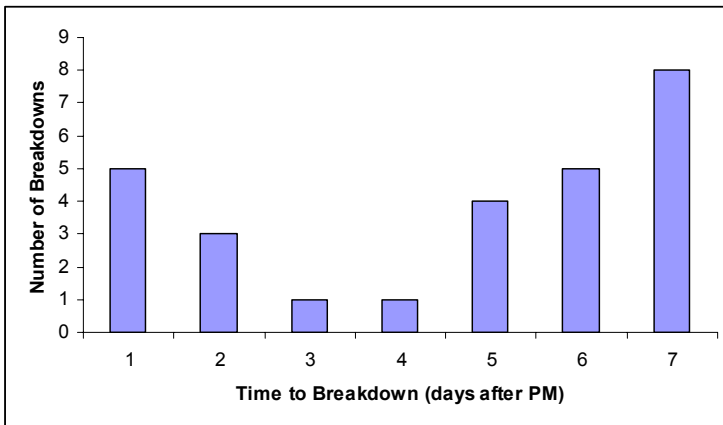


Figure 19.11. Illustrating the number of breakdowns to arrive after a PM

In the limiting case, the probability of detecting and removing j_i defects at inspection i (time iT) is still

$$P(N_s(iT) = j_i) = \frac{E[N_s(iT)]^{j_i} e^{-E[N_s(iT)]}}{j_i!} \quad (19.18)$$

The likelihood function for the refined intervals is now the product of the probability associated with observing the number of repairs at each inspection and the probability of observing the number of failures in each increment for all the intervals:

$$Likelihood = \prod_{i=1}^K \left(P(N_s(iT) = j_i) \prod_{b=1}^z P(N_f(I_b^i) = m_{ib}) \right) \quad (19.19)$$

As before, taking logarithms of the likelihood function reduces the complexity from an optimization perspective.

19.5 A Case Example

A copper works in northwest England has used the same extrusion press for over 30 years, and the plant is a key item in the works since 70% of its products will go through this press at some stage of their production. The machine comprises a 1700-ton oil-hydraulic extrusion press with one 1700-kW induction heater and completely mechanized gear for the supply of billets to the press and for the removal of the extruded products. The machine was operated for 15–18 hours a day (two shifts), 5 days a week, excluding holidays and maintenance down-time. Preventive maintenance (PM) had been carried out on this machine since 1993, which consisted of a thorough inspection of the machinery, along with any subsequent adjustments or repairs if the defects found could be rectified within the PM period. Any major defects which could not be rectified during the PM time were supposed to be dealt with during non-production hours. PM lasted about 2 h and is performed once a week at the beginning of each week.

Questions of concern are (1) whether PM is or could be effective for this machine; (2) whether the current PM period is the right choice, particularly, the one week PM interval which was based upon maintenance engineers' subjective judgement; (3) whether PM is efficient, *i.e.*, whether it can identify most defects present and reduce the number of failures caused by those defects.

In this case study, the delay time model introduced earlier was used to address the above questions. The first question can also be answered in part by comparing the total downtime per week under PM with the total downtime per week of the previous years without PM. A parallel study carried out by the company revealed that PM has lowered the total downtime. The proportion of downtime has reduced from 7.8% to 5.8%.

To establish the relationship between the downtime measure and the PM activities using the delay time concept, the first task is to estimate the parameters of the underlying delay time distribution from available data, and hence build a model to describe the failure and PM processes.

For a detailed description of the data and the parameters estimation process, see Christer *et al.* (1995) and here we briefly illustrate the parameter estimation process. In the original study, Christer *et al.* (1995), a number of different candidate delay time distributions were considered including exponential and Weibull distributions.

The chosen form for the delay time distribution is a mixed distribution consisting of an exponential distribution (scale parameter α) with a proportion P of defects having a delay time of 0. The cdf. is given by

$$F(h) = 1 - (1 - P)e^{-\alpha h}$$

An optimization algorithm is required for maximisation of the likelihood with respect to the parameters. The estimated values using Equation 19.19 are given in Table 19.1 with their associated coefficients of variation (CV).

Table 19.1. Estimated model parameters

	Rate of occurrence of defect	Probability of perfect inspection	Proportional of zero delay time of defects	Scale parameter
	$\hat{\lambda} = 1.3561$	$\hat{r} = 0.902$	$\hat{P} = 0.5546$	$\hat{\alpha} = 0.0178$
CV	0.0832	3.4956	0.4266	1.1572

Inserting the optimal parameter estimates into the log-likelihood function gives an ML value of 101.86.

It is noted that the delay time distribution is a mixture of an exponential distribution with a proportion P of defects with zero delay time. This has been confirmed by the data and the Akaike Information Criterion (AIC) (Baker and Wang, 1991).

The downtime model is the same as Equation 19.1, that is

$$D(T) = \frac{d_f E[N_f(T)] + d_s}{T + d_s}$$

and the expected number of failures is given by, Equation 19.9 with $i \rightarrow \infty$ for effectively, the system is in a steady state:

$$E[N_f(T)] = \lambda \left\{ (e^{-\alpha T} - 2 + e^{\alpha T}) \frac{1-r}{\alpha(e^{\beta T} - 1 + r)} (1-P) + T + \frac{1}{\alpha} (e^{-\alpha R} - 1)(1-P) \right\}$$

The mean down time per failure, d_f , and per PM, d_s , were obtained from history data of failures and PM. It was found that in this case the downtime per failure is a function of the interval between PM since the company did some experiments before finally moving to the weekly PM. This is reasonable since serious failures may be more likely to occur in longer PM intervals than in shorter ones. Based on a simple regression analysis, it turned out that

$$d_f = \begin{cases} 20 & \text{if } T \leq 7 \text{ days} \\ 17.67 + 0.33T & \text{if } 7 < T \leq 49 \\ 34 & \text{if } T > 49 \end{cases}$$

In the first half of 1993, the PM activity performed on the press used to occupy 2 h of production time, *i.e.*, it caused 2 h of downtime. Later on, as the technicians gained experience, and particularly when the management in the factory allowed early-morning access to the plant before production started, the downtime caused by PM reduced to 30 min. This subsequently decreased to zero because all the PM activity was scheduled and completed before production started. Therefore, d_s was set to be 120, 30 and 0 min respectively.

Now all the parameters are ready and the computing of Equation 19.1 is straightforward. A graphical output is shown in Figure 19.12.

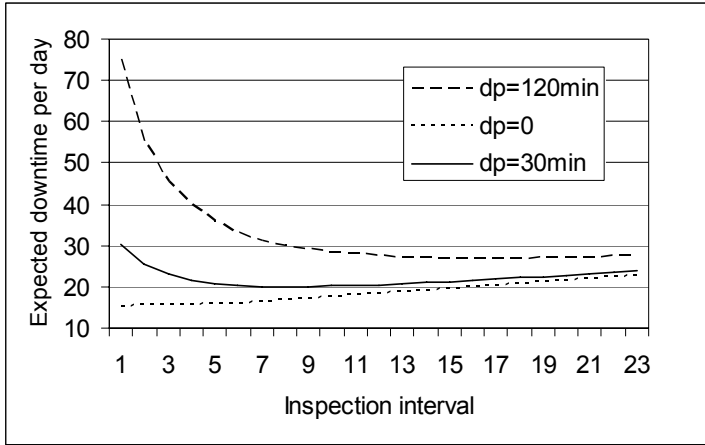


Figure 19.12. Expected downtime per day vs PM inspection intervals (days)

From Figure 19.12 it can be seen that a daily PM is the optimal choice if the downtime for PM is zero; also if the downtime due to PM is about 2 h, the optimal PM interval should be around 2–3 weeks. It is clear that if d_s is 30 min, then a weekly PM cycle is best. However, it is impractical to check the machine every day and occupy 2 h of the maintenance staff's time, due to limited manpower and the cost of overtime. The model confirmed that the company's current weekly PM inspection with about 30 min production downtime is the best option.

The downtime per press hour in 1992 when no PM was undertaken, and the downtime over periods of both weekly and daily PM policies in 1993, were available to use from production records. To compare model outputs with the observed downtime under various PM schemes, percentage downtime from production records and the model outputs are shown in Table 19.2.

Table 19.2. Downtime percentage

PM policy	(by production record)	(by model output)
No PM	5.47	5.53
1 week PM cycle	4.06	4.05
1 day PM cycle	2.45	1.85

It can be seen from Table 19.2 that, with the exception of the daily PM case, where the model underestimates the downtime actually observed, the model output gives values in close agreement with production records. It should be pointed out that the observed data from the daily PM was based upon only 1 month's operational data, while for the no PM and weekly PM approximately 6 months and 12 months of operational data were available, so we have less confidence on the recorded data from the daily PM case. Overall, the study confirms the validity of the modeling and inspires confidence that appropriate OR modeling can be used in supporting maintenance decisions.

19.6 Other Developments in DT Modeling and Future Research Directions

Several extensions have been made over the last decade to make the delay time model more realistic, but that increases the mathematical complexity as well.

This chapter has focused on complex system inspection modeling using the delay time concept, while researches have been done on single component units subject to a single failure mode based the DT concept (Baker and Wang, 1991, 1993; Wang and Christer 1997). If the arrival rate of defects was not constant, Christer and Wang (1995) addressed an NHPP non-perfect inspection delay time model of multiple component systems. In this case the constant inspection interval assumption cannot be held, and a recursive algorithm was developed in Wang and Christer (2003) to find the optimal non-constant intervals till final replacement. Christer *et al.* (1997) used an NHPP for modeling the rate of occurrence of defects in a case study of steel production plant where the initial rate of defect arrivals is higher just after a PM. Wang (2000) developed a model of nested inspections using the delay time concept. Wang and Jia (2007) reported the use of empirical Bayesian statistics in the estimation of delay time model parameters using subjective data, which overcame a number of problems in previous subjective delay time parameter estimation (Christer and Waller, 1984; Wang 1997). If the times of failures were available, Christer *et al.* (1998) developed an extension to the method introduced in this chapter to incorporate this additional information in DT model parameter estimation. If the downtime due to failures could not be ignored in the calculation of the expected number of failures during an inspection interval, Christer *et al.* (2000) addressed this problem and a refined method was proposed. Christer *et al.* (2001) compared the delay time model with an equivalent semi-Markov setting to explore the robustness of both modeling techniques to the Markov assumption. A recent case study on the use of DT modeling is reported in Akbarov *et al.* (2008) where a baking line was modeled and both objective and subjective data were used.

The future research on the DT modeling relies on the application areas, the data involved, and the objective function chosen. We consider that the following areas or problems are worthy of research using the delay time concept:

1. PM type of inspections. Inspections may consist of many activities and some of them are purely preventive types such as greasing, top-up oil, and cleaning, which may have no connection with defect identification. It is noted, however, that this type of PM may change the RATE of defect arrivals and therefore change the expected number of failures within an inspection interval. This problem has not been modeled in previous DT research, but it is a reality we have to face. An initial idea is to introduce another parameter in the RATE OF DEFECT ARRIVALS to model the effectiveness of such PM activities.
2. Multiple inspections scheme. This is again common in practice in that more than one inspection intervals of different scales or types are in place. Wang (2000) developed a DT model for nest inspections, but

the model is not generic, and can only be used for a specific type of problems.

3. Condition monitoring (CM) is becoming more popular in industry and offers abundant modeling opportunities with a large amount of data. With CM it may be possible to identify the initial point of a random defect at an earlier stage than that of using manual inspections, and it is possible that u becomes observable by CM. A pilot research has been carried out to investigate the use of the DT concept in condition based maintenance modeling (Wang, 2006).
4. Parameters estimation. This is still an on-going research since for each specific problem we may have to develop a tailor made approach. The empirical Bayesian approach outlined earlier is promising since it combines both subjective and objective data. It is noted, however, that the computation involved is intensive and, therefore, algorithms developments are required to speed up the process.

References

- Akbarov A, Wang W and Christer AH (2008), Problem identification in the frame of maintenance modeling: a case study. *I J Prod Res* 46(4): 1031–1046.
- Baker RD, and Wang W (1991) Estimating the delay time distribution of faults in repairable machinery from failure data. *IMA J Math Appl in Bus and Ind* 4: 259–282.
- Baker R and Wang W (1993) Developing and testing the delay time model. *J of Oper Res Soc* 44(4): 361–374.
- Christer, AH (1976) Innovative Decision Making. Proc. NATO conference on the role and effectiveness of theories of decision in practice, eds K.C Bowen and D.J. White, Hodder and Stoughton, pp 368–377
- Christer, AH (1999) Developments in delay time analysis for modeling plant maintenance. *J Oper Res Soc* 50: 1120–1137.
- Christer AH and Waller WM (1984) Delay time Models of Industrial Inspection Maintenance Problems. *J Oper Res Soc* 35: 401–406.
- Christer AH and Wang W (1995) A delay time based maintenance model of a multi-component system. *IMA Journal of Maths. Appl in Bus and Ind* 6: 205–222.
- Christer AH, Wang W, Baker RD, and Sharp J (1995) Modeling maintenance practice of production plant using the delay time concept. *IMA J Math Appl in Bus and Ind* 6: 67–83.
- Christer AH, Wang W, Sharp J, and Baker RD (1997) A stochastic modeling problem of high-tech steel production plant, in *Stochastic Modeling in Innovative Manufacturing, Lecture Notes in Economics and mathematical Systems*, (Eds. by A.H Christer, Shunji Osaki and L. C. Thomas), Springer, Berlin, pp 196–214.
- Christer AH, Wang W, Choi K and Sharp J (1998) The delay-time modeling of preventive maintenance of plant given limited PM data and selective repair at PM. *IMA J Math Appl in Bus and Ind* 9: 355–379.
- Christer AH Wang W and Lee C (2000) A data deficiency based parameter estimating problem and case study in delay time PM modeling. *Int J Prod Eco* 67: 63–76.
- Christer AH, Wang W, Choi K, and Schouten FA (2001) The robustness of the semi-Markov and delay time maintenance models to the Markov assumption. *IMA J Manag Math* 12: 75–88.

- Taylor HM, and Karlin S (1998) An introduction to stochastic modeling, 3rd Edn, Academic press, San Diego.
- Wang W (1997) Subjective estimation of the delay time distribution in maintenance modeling. *Eur J of Oper Res* 99: 516–529.
- Wang W (2000) A model of multiple nested inspections at different intervals. *Comput and Oper Res* 27: 539–558.
- Wang W (2006) Modeling the probability assessment of the system state using available condition information. *IMA J Manag Math* 17(3): 225–234.
- Wang W and Christer AH (1997) A modeling procedure to optimize component safety inspection over a finite time horizon. *Qual and Reliab Eng Int* 13(4): 217–224.
- Wang W and Christer AH (2003) Solution algorithms for a multi-component system inspection model. *Comput and Oper Res* 30: 190–193.
- Wang W and Jia X (2007) A Bayesian approach in delay time maintenance model parameters estimation using both subjective and objective data. *Qual Maint and Reliab Int* 23: 95–105.

Integrated E-maintenance and Intelligent Maintenance Systems

Jayantha P. Liyanage, Jay Lee, Christos Emmanouilidis, and Jun Ni

20.1 Introduction

Development and acquisition of technological capabilities has become one of the major strategic requirements to excel commercially today in various business sectors. Front-end innovative technologies in conjunction with mass globalization and outsourcing of industrial operations has created a fruitful environment for technology-based growth and excellence. Different industries are in search for various technological solutions in their continuous efforts to improve performance in different parts of their businesses. Advanced solutions are often found implemented within range of application areas varying from corporate information management to logistics planning and coordination activities. In this setting, both industrial assets and business processes have been subjected to a technology-driven change process.

As the industry began to pay more and more emphasis on quality, precision, task sensitivity, and product life cycle considerations, the automation solutions for production, manufacturing, and process plants are gradually taking central stage. This brought a major impact on the massive use of robot technology, electronics, advanced programming and mathematical modeling, that were earmarked for sophisticated technical solutions within the operational environments. Most of the complex and capital-intensive industrial plants and facilities, in particular, displayed the tendency to become fully or semi-automated, targeting various business benefits. Some of the core technologies were seen put into very practical and productive use during this period in production, manufacturing, and process environments. Use of such technologies for operational purposes still continues and apparently grows towards more advanced level of applications for relatively complex usage.

The growth of information and communication technologies (ICTs) has certainly brought a new dimension to the industrial plant or the facility environment today. This is not only in terms of abilities for creating repositories of gigabytes of data in comprehensive *Enterprise Resource Planning* (ERP) systems,

but also with respect to effective and efficient management of daily plant operations and maintenance activities. ICT is in fact a principal landmark in this setting today, and a major contributor to the current level of sophistication in the use of advanced technical solutions to resolve plant or facility related problems. With the parallel advancement in instrumentation technologies, analytical software and mathematical modeling, the industry has been presented with substantial potential to implement innovative solutions to improve operations and maintenance (O&M) practice. This has brought much optimism to different businesses, which still rely much on the conventional O&M practices, pushing those industrial sectors to exploit numerous opportunities to reduce commercial risks associated with plant operations.

Technical condition and safety integrity of plants or assets in operation are defining factors for risk mitigation and value creation. Formally, the technical condition can explicitly or implicitly be expressed by means of different terms including reliability, availability on demand, downtime (or uptime), history of failure, actual capacity utilization, failure frequency, and scale of losses. Obviously, the behavior of systems and equipment under a given operational setting, their functional characteristics, and the technical faults and failures, actively contribute in defining technical conditions of operating plants or assets. It implies that the ability of the operator to identify systems or equipment malfunctions prior to any unwanted event or an incident is very important part of risk mitigation and value creation efforts. In principle, such ability relies much on the technical data obtained from technical systems and equipment, and the decision support setting of the operator. Proper instrumentation of critical systems and equipment plays a vital role in the acquisition of necessary technical data, while the support of analytical software with embedded mathematical models is crucial for the decision making process. This explanation in fact presents the very basics of the O&M intervention process that aims at retaining or restoring systems or equipment in a particular condition so that the plant or the asset complies with specific level of performance (see Figure 20.1). The technical instruments in use and the analytical software and tools provide necessary engineering basis to monitor the condition of systems and equipment of any given asset. Results from this condition monitoring process are the inputs to the decision platforms and processes of the plant or the asset operator in making diagnostic or prognostic decisions. If a fault or a failure is imminent, then necessary work orders are issued for the O&M crew.

This O&M intervention process illustrates the very basic concept behind *condition-based maintenance (CBM)* practice. The wide spreading concepts of e-maintenance and intelligent maintenance systems can exploit the availability of a CBM platform and take the form of advanced applications employing modern ICT, robust technical infrastructures, and sophisticated electronic gadgets and data acquisition technologies.

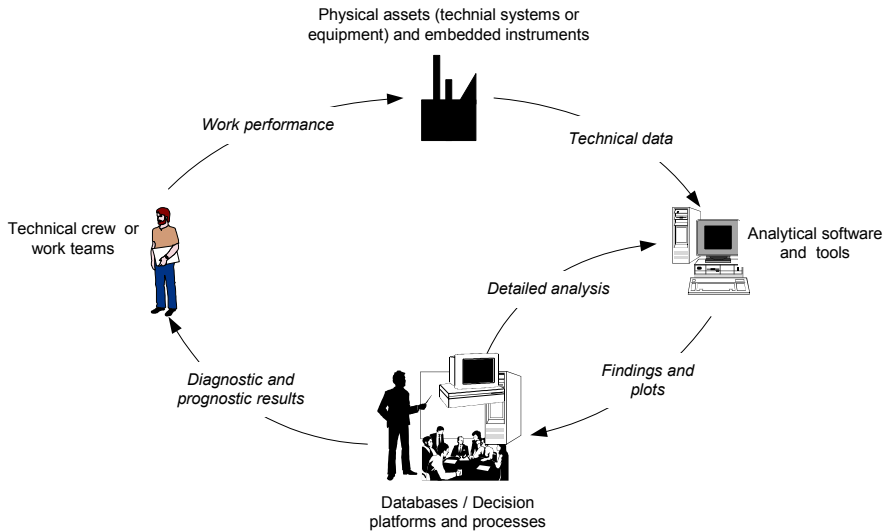


Figure 20.1. Basic O&M intervention process to retain or to restore technical systems and equipment of an industrial asset in an acceptable technical condition

20.2 Condition-based Maintenance Technology and the State of Development

As the commercial implications of technical systems' malfunctions and non-availability become more apparent, industrial organizations have begun to resort to novel means to address technical systems' performance challenges. Notably, most machine field services today depend on sensor-driven management systems that provide alerts, alarms, and indicators. The moment the alarm sounds, in most cases it's already too late to prevent the failure. Therefore, most machine maintenance today is either purely reactive (fixing or replacing equipment after it fails) or blindly proactive, assuming a certain level of performance degradation, with no input from the machinery itself, and servicing equipment on a routine schedule whether service is actually needed or not. Both scenarios could be extremely wasteful.

Substantial research efforts have been devoted to machinery fault diagnostics in reducing downtime. The preventive maintenance (PM) scheme is time-based without considering the current health state of the machine, and hence leads to unnecessary maintenance. A predictive maintenance (PdM) scheme appeared subsequently, and was presented as a maintenance scheme to provide sufficient warning of an impending failure on a particular piece of equipment, allowing that equipment to be maintained only when there is objective evidence of an impending failure. Condition based maintenance (CBM) is currently a popular scheme of PdM. CBM methods and practices have been continuously improved in recent decades. Sensor fusion techniques are now commonly in use due to the inherent

superiority in taking advantage of information retrieval from multiple sensors (Hansen *et al.* 1994; Reichard *et al.* 2000; Roemer *et al.* 2001). A variety of techniques in vibration, temperature, acoustic emissions, ultrasonic, oil debris, lubricant condition, chip detectors, and time/stress analyses have received considerable attention. For example, vibration signature analysis, oil analysis and acoustic emissions, because of their excellent capability of describing machine performance, have been successfully employed for prognostics for a long time (Kemerait, 1987; Wilson *et al.* 1999, Goodenow *et al.* 2000). Current prognostic approaches can be classified into three basic groups namely:

- *Model-based approach*: requires detailed knowledge of the physical relationships between, and characteristics of, all related components in a system. It is a quantitative model used to identify and evaluate the difference between the actual operating state determined from measurements, and the expected operating state derived from the values of the characteristics obtained from the physical model, see for instance Bunday (1991) who presented the theory and methodology of obtaining reliability indices from historical data. However, it is usually prohibitive to use the model-based approach since relationships and characteristics of all related components in a system and its environment are often too complicated to build a model with acceptable accuracy. Furthermore, the values of some process parameters/factors may not be readily available. A poor model leads to poor judgment.
- *Data-driven approach*: requires a large amount of history data, representing both normal and “faulty” operation. It uses no prior knowledge of the process, but instead derives behavioral models only from measurement data from the process itself. Pattern recognition techniques are widely used in this approach. General knowledge of the process can be used to interpret results from data analysis, based on which qualitative methods such as fuzzy logic, and artificial intelligence methods can be used for decision making to enable fault prevention.
- *Hybrid approach*: fuses the model-based information and sensor-based information and takes advantage of both model-driven and data-driven approaches through which a more reliable and accurate prognostic results can be generated (Hansen *et al.* 1994). Garga (2001) introduced a hybrid reasoning method for prognostics, which integrated explicit domain knowledge and machinery data. In this approach, a feed-forward neural network was trained using explicit domain knowledge to get a parsimonious representation of the domain.

However, a major breakthrough has not been made since. Existing prognostic methods are application or equipment specific. For instance, the development of neural networks has added new dimensions to solving existing problems in conducting prognostics of a centrifugal pump case (Liang *et al.* 1988). A comparison of the results using the signal identification technique shows various merits of employing neural nets including the ability to handle multivariate wear parameters in a much shorter time. A polynomial neural network was conducted in fault detection, isolation, and estimation for a helicopter transmission prognostic application (Parker *et al.* 1993). Ray and Tangirala (1996) built a stochastic model

of fatigue crack dynamics in mechanical structures to predict remaining service time. Fuzzy logicbased neural networks have been used to predict paper web breakage in a paper mill (Bonissone, 1995) and the failure of a tensioned steel band with seeded crack growth (Swanson, 2001). Neurofuzzy and probabilistic neural network techniques have been employed for novelty detection and diagnostics of machinery, such as gearboxes and machine cutting tools (Emmanouilidis *et al.* 1998, 2006), while evolutionary multiobjective algorithms have been employed for selecting combinations of features for building diagnostic models (Emmanouilidis 2002). Yet another prognostic application presented an integrated system in which a dynamically linked ellipsoidal basis function neural network was coupled with an automated rule extractor to develop a tree-structured rule set which closely approximates the classification of the neural network (Brotherton *et al.* 1999). That method allowed assessment of trending from the nominal class to each of the identified fault classes, which means quantitative prognostics were built into the network functionality. Vachtsevanos and Wang (2001) gave an overview of different CBM algorithms and suggested a method to compare their performance for a specific application.

20.3 Integrated E-maintenance Solutions and Current Status

As mentioned earlier, *condition-based maintenance* (CBM) by definition concerns making decisions and performing necessary maintenance tasks based on the detection and monitoring of selected equipment parameters, the interpretation of readings, the reporting of deterioration, and the vital warnings of impending failure (Stoneham, 1998). In general, CBM can be based on both embedded and/or portable techniques leading to online or offline monitoring capabilities (Figure 20.2). The actual practice can be based on different measurement and detection methods based on the nature of the preferred technical parameter under surveillance and the operational setting. Techniques such as vibration analysis, acoustic emission, thermography, and lub-oil analysis have come in to common use over the last few years, together with various non-destructive techniques (NDT) such as visual inspection, magnetic particle inspection, and eddy current methods.

The industrial practice gradually showed an interest in adapting condition monitoring as a strategic tool to resolve some major challenges in various plants, facilities, and industrial settings. Subsequently, several condition monitoring solutions appeared in the market, both as ‘off-the-shelf solutions’ or in the form of customizable solutions. In fact it is the continuous development of CBM expertise coupled with data acquisition and presentation software that has laid a solid foundation for further development of technology-based maintenance efforts leading the path towards more advanced diagnostics and prognostics solutions.

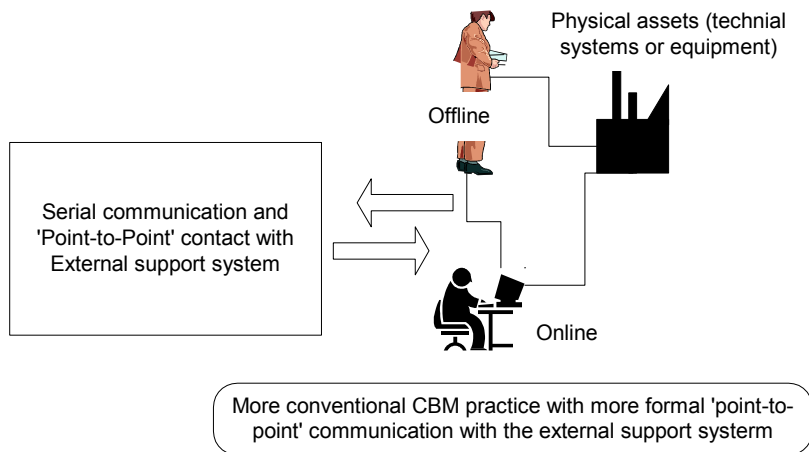


Figure 20.2. Industrial CBM practice in the more conventional form

Prognostic information, obtained through the intelligence embedded into the manufacturing process or equipment, can also be used to improve manufacturing and maintenance operations in order to increase process reliability and improve product quality. For instance, the ability to increase reliability of manufacturing facilities using the awareness of the deterioration levels of manufacturing equipment has been demonstrated through an example of improving robot reliability (Yamada and Takata, 2002). Moreover, a life cycle unit (LCU) (Seliger *et al.* 2002) was proposed to collect usage information about key product components, enabling one to assess product reusability and facilitating the reuse of products that have significant remaining useful life.

In condition monitoring it is simply not sufficient to base decisions on single-instance measurements. Information should represent a trend, not just a status. If machine health degradation can be monitored and degradation rate be predicted thereafter, maintenance actions can be taken when necessary (not too early or too late) before unacceptable levels of machine performance occur. This generates a critical need for prognostic capabilities to identify leading indicators of failure for accurate assessment of product damage significantly prior to appearance of any macro-indicators of the initiated damage. In addition to the need for prognostic capabilities, the highly dynamic nature of maintenance-related decision making requires one to strategically and intelligently utilize modern computing and communication technologies and coordinate them within security and communication bandwidth limitations, and also cost-effectively with respect to maintenance goals, in order to obtain maximal system level benefits.

Current development trends clearly indicate that the e-maintenance practice has taken the traditional CBM methodology to an advanced level of industrial application. This advancement has mostly been made possible by the rapid development of ICTs and network-based information and communication infrastructures. It implies that the success of the concept of e-maintenance largely rests on the data exchange and information sharing capabilities, and the ready

access to remotely located knowledge or competence pools to get expert assistance to solve technical mal-functions of the plant. In this context, e-maintenance can be defined as:

A concept based on an extended application of CBM where the technical condition of systems and equipment can remotely and jointly be monitored through active sharing of technical data and expertise between geographically dispersed locations enhancing diagnostic and prognostic capabilities by means of advanced information and communication technologies and networks so that well-coordinated decisions and actions can be taken through an organizational and service network to achieve near-zero down time performance.

As opposed to more frequently used local-area networks (LANs) to establish the connection between the machine and the technical expert, e-maintenance practice need solutions based on wide-area networks (WANs) and even web-based solutions. In an e-maintenance setting, the communication and exchange process takes place electronically within an authorized network of experts (for instance involving CBM experts, planning engineers, spare parts, and logistics personnel) simultaneously, even incorporating data filtering and semantic technologies (Figure 20.3). This is far beyond the formal ‘one-to-one’ connection setting of the conventional CBM practice that subsequently involves *serial* tasks to be performed by different technical groups.

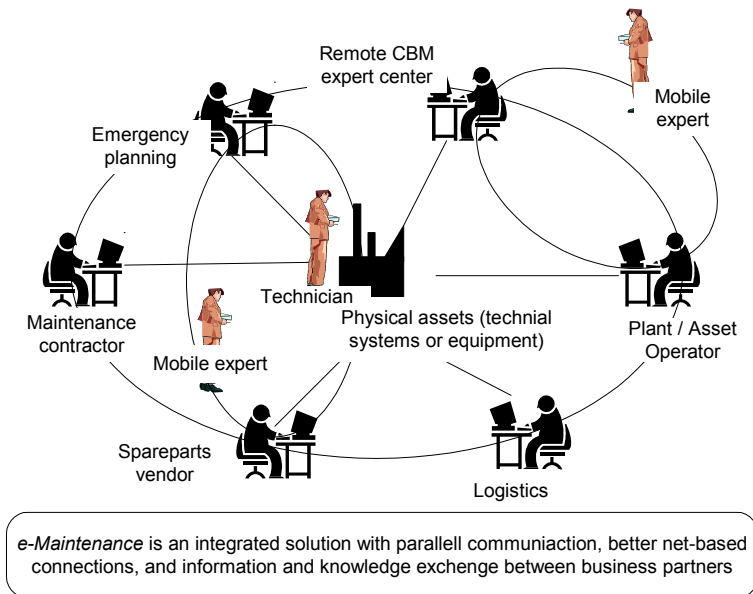


Figure 20.3. E-maintenance provides an integrated solution to manage the technical condition of industrial plants

As it appears, e-maintenance is by far an integrated approach to solve technical problems of systems and equipment of industrial plants through a collective effort. Even though e-maintenance has not yet achieved its full-blown engineering maturity, the rapid developments in sensor technology, video conferencing facilities, web-based data exchange and communication platforms, as well as portable and mobile technologies are contributing much to the continuous development of the practice.

Both the industrial and the academic environments, over the last few years, have given much attention to e-maintenance. This can largely be attributable to the growing concerns on the rising plant (or asset) operating costs, competence or knowledge gaps, and also the trend towards relying more and more on data-dependent decision support systems. Seemingly, e-maintenance gradually earns the recognition today as maintenance practice with considerable payback potential to achieve near-zero-downtime performance of systems and equipment cost-effectively.

As the condition monitoring applications are gradually growing in maturity, R&D activities by various expert groups have contributed further towards advanced and innovative solutions. Some of the current developments include, for instance, neural networks, expert systems, fuzzy logic, genetic algorithms, multi-agent platforms and case based reasoning, *etc.* (see for instance Liang *et al.* 1988; Yager and Zadeh, 1992; Jantunen *et al.* 1996; Lee, 1996; Chande and Tokekar, 1998; Sanz-Bobi and Toribio, 1999; Yang *et al.* 2000; Garga, 2001; Garcia and Sanz-Bobi, 2002; Marceguerra *et al.* 1998, 2006, Zio *et al.* 2002; Yu *et al.* 2003; Palluat *et al.* 2006). Moreover, the growth of R&D activities has resulted in introduction of novel application concepts and platforms such as PROTEUS (Bangemann *et al.* 2006), EXAKT (Jardine *et al.* 1998), Watchdog Agents[®] (Djurdjanovic *et al.* 2003), SIMAP (Garcia *et al.* 2006), *etc.*

The more recent focus appears to be on comprehensive technical solutions introducing what is termed Intelligent maintenance systems (IMS) (Sanz-Bobi *et al.* 2002; Iung, 2003; Lee, 2004; Moore and Starr, 2006). In principle, IMS constitutes more robust and comprehensive technical solutions where data acquisition, processing and interpretation, and decision support components are integrated. This has lead to some novel engineering tools such as 'Watchdog Agents' that includes a series of toolboxes for signal processing and system performance evaluation (Lee *et al.*, 2006). The toolboxes include signal processing and feature extraction tools such as Fourier analysis, time-frequency distribution, wavelet packet analysis and time series models, while performance evaluation tools incorporate fuzzy logic, match matrix, neural network and other advanced algorithms (Jardine *et al.* 2006).

Condition monitoring and e-maintenance solutions have been widely acknowledged as the cost-effective maintenance practice for various industries. Notably, there are some variations in the type of solutions preferred or suitable and the nature of the practical applications and use depending on the commercial challenges and the available technical infrastructure for plant operations.

20.4 Technical Framework for E-maintenance

As mentioned earlier, e-maintenance presents an integrated framework for industrial applications. The success with which e-maintenance is used for commercial benefits relies largely on three-fold issues, namely:

Application technology in use;
Business-to-business organizational solutions; and
Work performance by role and responsibility.

The type of application technology used by an industrial plant can vary depending on a suitable data management strategy and the actual data handling practice needed. Such technology can be put into practical use for various purposes within e-maintenance, particularly for data acquisition, data interpretation and visualization, and data/information and knowledge exchange and communication. The data acquisition tasks within a plant rely in principle on sensor technologies and instrumentation techniques. In a more macro scale, such as for tracking and monitoring of logistics, other location-based techniques, such as radio frequency based identification (RFID) and global positioning (GPS), can also be used for indirect and direct localization respectively. Such macro scale applications apply largely to mobile assets such as cargo vessels, drilling rigs and military ships. The interpretation and visualization needs, on the other hand, have to be addressed based on various signal handling, analytical and presentation software products. Often, specific data mining techniques would also have to be configured so that the analysis can be carried out based on both currently reported data as well as historical data available in corporate databases. Apart from the above two technological types, the data/information and knowledge exchange and communication can constitute a range of functional requirements within an e-maintenance environment. Proper and more effective use of e-maintenance practice for commercial advantage calls for advanced and reliable wide-area network (WAN) solutions. They need to be able to provide integrated solutions through a dedicated and high secure ICT infrastructure and/or through world wide web-based (www-based) authorized access. Current applications in this context mostly appear to rely on web-based solutions. Recent developments within mobile and wireless technologies have offered novel and innovative solutions in this context, enabling such devices as personal digital assistants (PDAs) and smart phones, to play a specific role within e-maintenance settings not only in terms of enhanced remote communication but also in exchanging plant or equipment critical data.

Clearly, advancement in data management technologies coupled with the developments within ICT infrastructures or network solutions introduce a new organizational form to operationalize e-maintenance solutions. The ideal organizational form is such that different business partners and technical experts remain connected within a common network solution so that they can interact by exchanging data and expertise regardless of the geographical location. This setting largely establishes a virtual organization involving various business-to-business solutions between cooperating organizations. A major development in this context, is what is called '*remote support centers*'. These are externally located support centers (*i.e.*, in a different location than where the plant or the facility is based) with necessary technologies and technical knowledge to constantly monitor and to

provide expertise or to take specific actions whenever necessary depending on systems performance or equipment condition. If a plant or a facility has its own 'control room', then *remote support centers* need be connected to both the organizational set-up of the operator of the plant and the *control room* located in the plant itself.

Roles and responsibilities assigned to different business partners and technical experts is an important component of the operational organization to keep the e-maintenance solution fully and reliably functional. The underlying roles and responsibilities may relate to activities such as logistics support and handling, vibration monitoring and condition assessment of specific production critical equipment, emergency response, remote instructions (or '*tele-instructions*') for specific technical tasks that require specific competence (e.g., turbine maintenance), coordinated troubleshooting with plant personnel, etc. The clarity of assignments has a direct influence both on the specification of decisions to be taken, and tasks and activities to be performed by internal and/or external competence sources. Different software products and corporate IT tools play a major role here providing the necessary technical basis for data management, work coordination and execution, and reporting and communication.

In fact all the three areas; *application technology*, *business-to-business organizational solutions*, and *work performance*, are truly inter-dependent within an e-maintenance setting. They are the cornerstones for a range of integrated e-maintenance solutions. However, the development and implementation of such comprehensive solutions needs to be a step-wise process that leads to the systematic development and integration of various engineering and managerial components of the e-maintenance system. In an ideal situation it should be the type of maintenance programs and the underlying tasks and activities that need to provide the foundation to build up an appropriate e-maintenance solution. For instance, the type and the nature of maintenance related tasks or activities, the volume of work involved and the need for specific technical competencies, all have major impact on the choice of external expertise and the assignment of specific roles and responsibilities. This leads towards necessary business-to-business solutions where a number of different organizations are involved in managing the condition of systems and equipment of a plant. The ICT and other technological solutions are configured or put into use to facilitate work that is performed by different technical experts, both internal and external. This is illustrated in Figure 20.4.

The development and use of comprehensive e-maintenance solutions largely depends on the tight coupling between *data*, *knowledge*, and *maintenance tasks*. The technical framework to operationalize a proper diagnostic and prognostic process through such an integrated approach presupposes a number of important functions, inclusive of:

- Data generation (coupled with data acquisition);
- Data coding/decoding and presentation;
- Data interpretation and exchange;
- Data analysis, mining, and simulation;
- Communication and knowledge exchange;
- Preparation of results and report generation;

- Joint decision making;
- Coordinated work planning;
- Developing emergency response;
- Response specification in the event of a potential fault or a failure; and
- Instructions and work execution supporting.

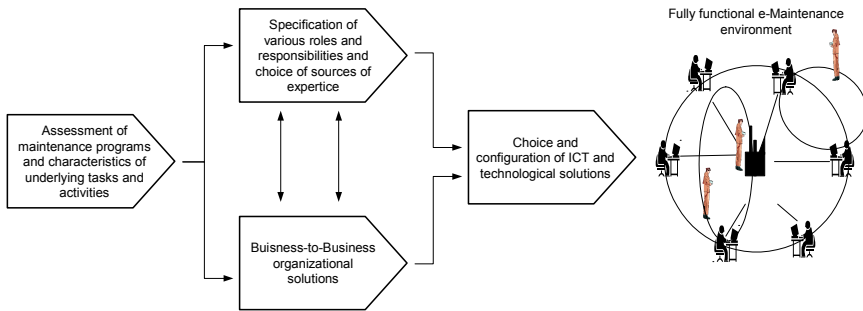


Figure 20.4. Establishing an integrated e-maintenance solution

It also implies that comprehensive and successful e-maintenance solutions demand a synergy between various competence groups from different fields of expertise. The strategies to establish the necessary competence group cooperation can largely be affected by the technical complexity of the facility by design (e.g., offshore oil and gas production assets in comparison to a fully automated automobile plant), systems or equipment ageing process, existing equipment maintenance contracts, access to ICT networks, varying production or process conditions, and inherent process complexities.

The underlying technical framework can be seen as built on three important functional windows, namely;

- Data generation and presentation window;
- Centralized ICT window; and
- Interpretation, decision making, and work planning window.

The '*data generation and presentation window*' involves the process of acquiring and sending out data to relevant databases either automatically or manually. The '*centralized ICT window*' facilitates the storage and distribution of received data as the data/information hub. In relatively large systems, this window is under the administration of an Internet/network provider. An example of such a case is the '*Secure Oil Information Link*' (so-called *SOIL*) network available in the North Sea for oil & gas exploration and production industry, operated and administrated by the network service provider *OilCamp*. In such large systems, special precautions are formally taken to build in the system specific reliability and security features. The access needs to be pre-authorized and in more advance cases logical filters can be built in to avoid exchange of repository of data that has no specific meaning or relevance to designated roles and responsibilities of a given industrial partner. The bandwidth of such *ICT windows* has a direct impact on the

exchange traffic, where data stream flows require higher bandwidths to accommodate a smooth and reliable exchange process. The ‘*interpretation, decision making, and work planning window*’, on the other hand, consists of the collaborative plant support system involving diagnostics, prognostics, decision making, and activity planning tasks. The industrial partners can gain access either to the *central ICT window* through local area networks (LANs), fiber-optic, satellite or different wide area network solutions (WANs), or resorting to web-based solutions through IP-VPN or ADSL. In addition, establishing remote wireless access has also become possible today with the use of mobile technologies and wireless application solutions. This is illustrated in Figure 20.5.

No specific set of technological solutions can be said to define integrated intelligent e-maintenance. Rather, various techniques such as fuzzy logic, artificial intelligence, neural network, genetic programming, logical reasoning, expert system, *etc.*, can certainly be useful in deriving different solutions under different conditions. Around the globe, leading research and development (R&D) institutes and eminent scholars invest considerable resources in the development of comprehensive solutions for application in a variety of industrial assets.

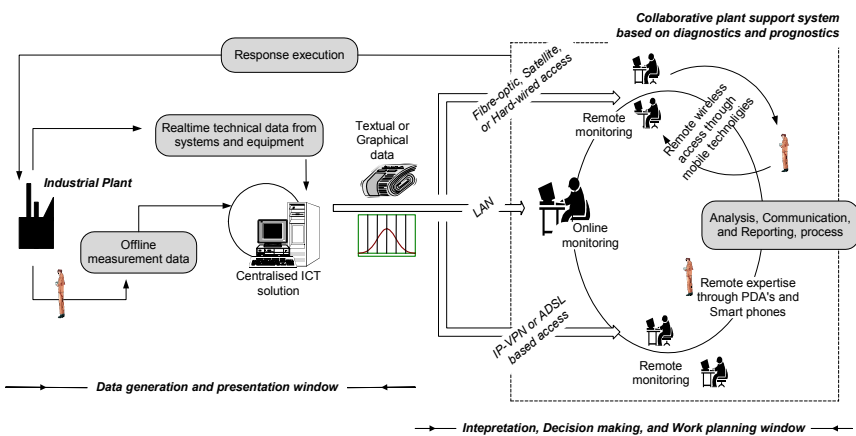


Figure 20.5. Generic technical framework for integrated e-maintenance solutions

To address the unmet needs and to overcome the new set of challenges emanated after “fail and fix” type diagnostic systems, the National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) on Intelligent Maintenance Systems (IMS) in the USA has taken specific steps to develop necessary technologies for development of “predict and prevent” prognostic methodologies. The emphasis has been on transforming on-line monitoring data as well as maintenance event data to prognostic health information to enable products and systems to achieve and sustain near-zero breakdown performance for improved productivity and asset utilization. The resulted application solution is discussed under the section ‘*Watchdog Agent-Based Intelligent Maintenance Systems*’.

20.5 Watchdog Agent-based Intelligent Maintenance Systems

Today most state-of-the-art manufacturing, mining, farming, and service machines (e.g., elevators) are actually quite “smart” in themselves. Many sophisticated sensors and computerized components are capable of delivering data concerning a machine’s status and performance. The problem is that little or no practical use is made of most of this data. We have the devices, but we do not have a continuous and seamless flow of information throughout entire processes. Sometimes this is because the available data is not rendered in a useable, or instantly understandable, form. More often, no infrastructure exists for delivering the data over a network, or for managing and analyzing the data, even if the devices were networked.

Watchdog Agent-Based Real-time Remote Machinery Prognostics and Health Management (R²M-PHM) system has been recently developed by the IMS Center. It focuses on developing innovative prognostics algorithms and tools, as well as remote and embedded predictive maintenance technologies to predict and prevent machine failures, as illustrated in Figure 20.6.

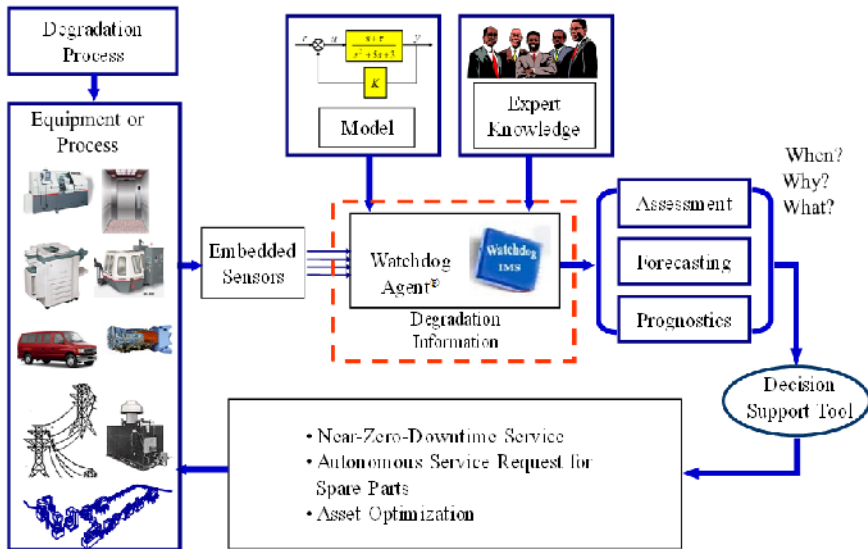


Figure 20.6. Key focus and elements of the intelligent maintenance systems

20.5.1 R²M-PHM Platform

A generic and scalable prognostics framework was presented by Su *et al.* (1999) to integrate with embedded diagnostics to provide “total health management” capability. A reconfigurable and scalable Watchdog Agent-based R²M-PHM platform is being developed by the IMS Center, which expands the well-known Open System Architecture for Condition-Based Maintenance (OSA-CBM) standard (Thurston and Lebold, 2001) by including real-time remote machinery diagnosis and prognosis systems and embedded Watchdog Agent technology. As

illustrated in Figure 20.7, the Watchdog Agent (hardware and software) is embedded onto machines to convert multi-sensory data to machine health information. The extracted information is managed and transferred through wireless internet or a satellite communication network, and autonomously trigger service and order spare parts.

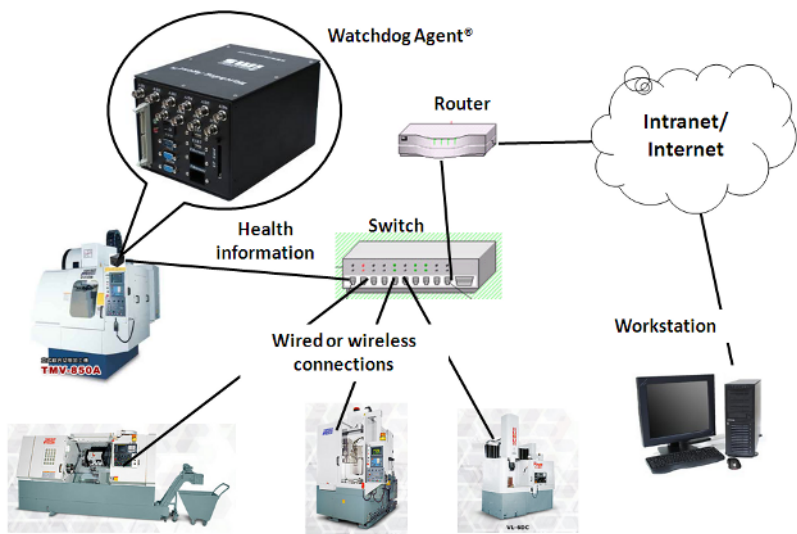


Figure 20.7. Illustration of IMS real-time remote machinery diagnosis and prognosis system

20.5.2 System Architecture

The system architecture of the Watchdog Agent-based R²M-PHM platform is shown in Figure 20.8.

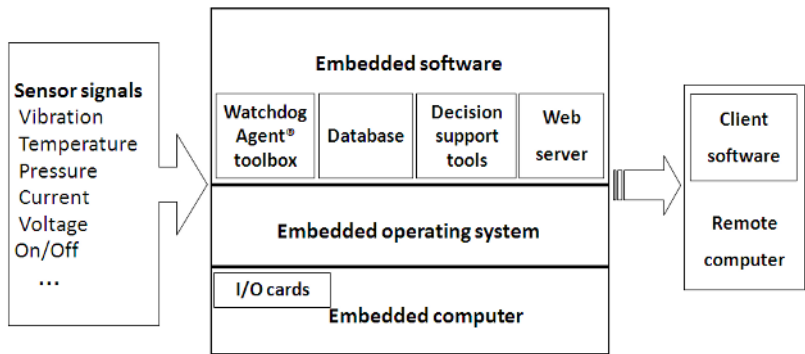


Figure 20.8. System architecture of a reconfigurable Watchdog Agent

In most products or systems, different sensors measure different aspects of the same physical phenomena. For example, sensor signals, such as vibration, temperature and pressure, are collected. In much the way that human “stereo” vision gives us depth perception, or multiple 2-D perspectives can be combined into a 3-D view, the IMS Center is working on software to “fuse” available data to form a more useable, holistic “image” of the actual state of a machine’s performance behavior. A “digital doctor” inspired by biological perceptual systems and machine psychology theory, the Watchdog Agent, consists of embedded computational prognostic algorithms and a software toolbox for predicting degradation of devices and systems. It is being built to be extensible and adaptable to most real-world machine situations. The health related information is saved to the database. The diagnostic and prognostic outputs of the Watchdog Agent, which is mounted on all the machinery of interest, can then be fed into the decision support tools. Decision support tools help the operation personnel balance and optimize their resources, when one or more machines are likely to fail, by constantly looking ahead. For example, if a production line has three processes A, B, and C, such that A has one machine, B has three machines, and C has one machine, what would we do if we could anticipate that one of the machines at station B is not behaving normally. Perhaps we’d arrange a staging area for output from A, or perhaps we’d ramp up production on the other two machines at station B. Whatever the case, we’d be making our decision before experiencing the impending breakdown. These tools are critical to maintenance and process personnel, enabling them to stay ahead of the game, balancing limited resources with constant change in demand. Decision support tools also helps minimize losses in productivity caused by downtime, and helps production and logistics managers optimize their maintenance schedule to minimize downtime costs. The lean and necessary information for maintenance can then be determined and become accessible through built-in web services.

The rapid development of web-enabled and cyber-infrastructure technologies are important enablers for remote monitoring and prognostics. One of the major barriers is that most manufacturers adopt proprietary communication protocols, which leads to difficulties in connecting diverse machines and products. Currently, the IMS Center is developing a web-enabled remote monitoring Device-to-Business (D2B)TM platform for remote monitoring and prognostics of diversified products and systems. A system methodology and infotronics platform has been developed that enables the transformation of product condition data into more a useful health information format for remote and network-enabled prognostics applications. The MIMOSA (Maintenance Information Management Open System Architecture) organization has adopted the IMS infotronic platform as one of its standard platforms and will use an IMS test-bed to demonstrate MIMOSA standards in its future activities. As shown in Figure 20.9, the IMS infotronics platform includes the Watchdog Agent toolbox (which contains adaptive algorithms for different situations and applications), decision support tools, data storage, and D2BTM (Device-to-Business) system level connectivity. The Watchdog Agent toolbox includes signal processing, feature extraction, performance assessment, autonomous learning, prediction and prognostics functions. The lean and necessary information for maintenance from decision

support tools can then be determined and sent out through D2B™ system level connectivity to remote workstations or computers.

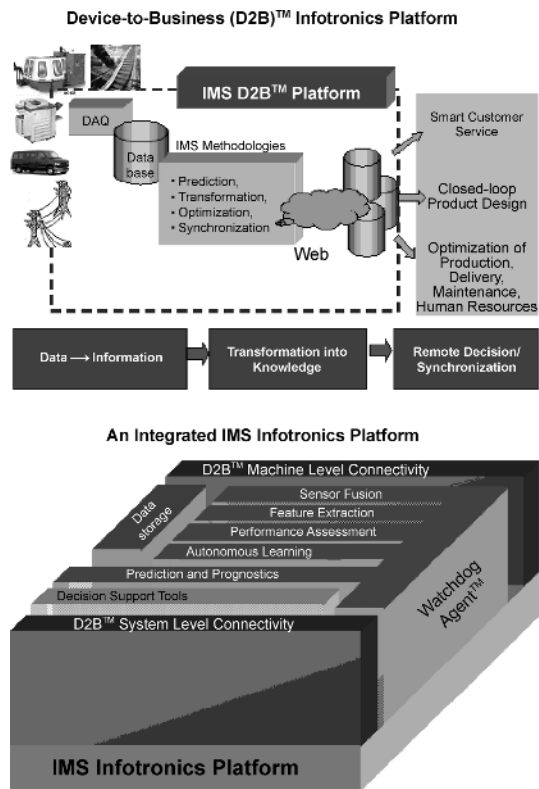


Figure 20.9. Integrated infotonics platform

20.5.3 Toolbox for Multi-sensor Performance Assessment and Prognostics

The Watchdog Agent toolbox, with autonomic computing capabilities, is able to convert critical performance degradation data into health features and quantitatively assess their confidence value to predict further trends so that proactive actions can be taken before potential failures occur. Figure 20.10 illustrates one of the developed enabling prognostics tools that can assess and predict the performance degradation of products, machines and complex systems.

The Watchdog Agent toolbox enables one to assess quantitatively and predict performance degradation levels of key product components, and to determine the root causes of failure (Casoetto *et al.* 2003; Djurdjanovic *et al.* 2000; Lee, 1995, 1996), thus making it possible to realize physically closed-loop product life cycle monitoring and management. The Watchdog Agent consists of embedded computational prognostic algorithms and a software toolbox for predicting degradation of devices and systems. Degradation assessment is conducted after the critical properties of a process or machine are identified and measured by sensors.

It is expected that the degradation process will alter the sensor readings that are being fed into the Watchdog Agent, and thus enable it to assess and quantify the degradation by quantitatively describing the corresponding change in sensor signatures. In addition, a model of the process or piece of equipment that is being considered, or available application specific knowledge can be used to aid the degradation process description, provided that such a model and/or such knowledge exist. The prognostic function is realized through trending and statistical modeling of the observed process performance signatures and/or model parameters.

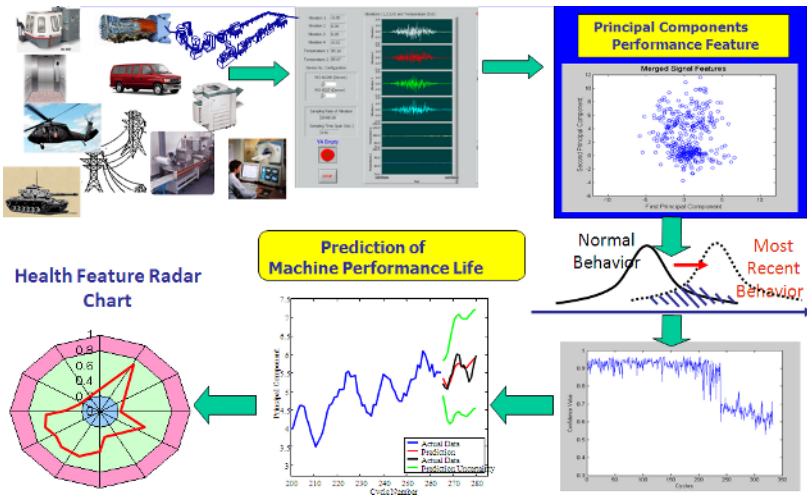


Figure 20.10. IMS innovation in advanced prognostics

In order to facilitate the use of Watchdog Agent in a wide variety of applications, with various requirements and limitations regarding the character of signals, knowledge of the mechanism of the deterioration process, loading conditions, deterioration conditions, feature dimensionality, trade-off between computing time and prediction accuracy, easiness of result interpretation, available processing power, memory and storage capabilities, as well as user preferences, the performance assessment module of the Watchdog Agent has been realized in the form of a modular, open architecture toolbox. The toolbox consists of different prognostics tools, including neural network-based, time-series based, wavelet-based and hybrid joint time–frequency methods, for predicting the degradation or performance loss on devices, process, and systems. The open architecture of the toolbox allows one to add new solutions easily to the performance assessment modules as well as to interchange different tools easily, depending on the application needs. To enable rapid deployment, a Quality Function Deployment (QFD) based selection method had been developed to provide a general suggestion to aid in tool selection; this is especially critical for those industry users who have little knowledge about those algorithms. The current tools employed in the signal processing and feature extraction, performance assessment, diagnostics and

prognostics modules of Watchdog Agent functionality are summarized in Figure 20.11.

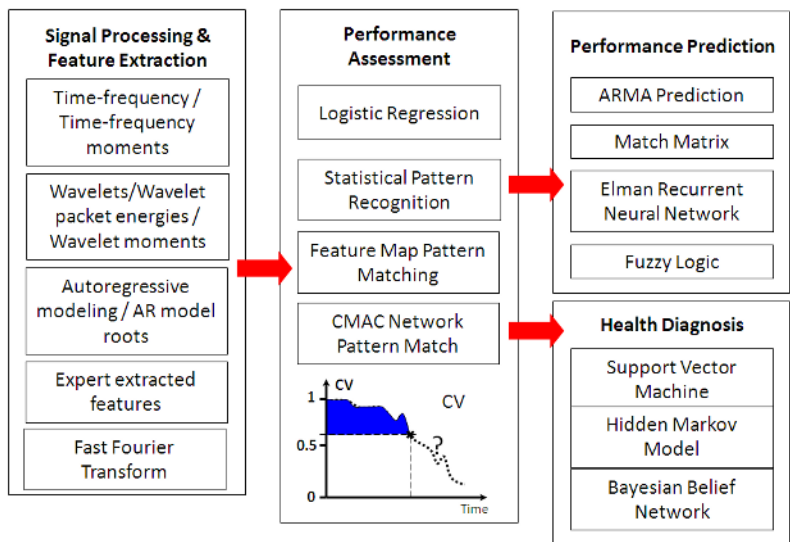


Figure 20.11. Watchdog Agent prognostics toolbox

Each of these modules is realized in several different ways to facilitate the use of the Watchdog Agent in a wide variety of products and applications.

Signal Processing and Feature Extraction Module

The signal processing module transforms multiple sensor signals into domains that are the most informative of a product's performance. Time-series analysis (Pandit and Wu, 1993) or frequency domain analysis (Marple, 1987) can be used to process stationary signals (signals with time invariant frequency content), while wavelet (Burrus *et al.* 1998), or joint time-frequency analysis (Cohen, 1995; Djurdjanovic *et al.* 2002) could be used to describe non-stationary signals (signals with time-varying frequency content). Most real life signals, such as speech, music, machine tool vibration, and acoustic emission are non-stationary signals, which place a strong emphasis on the need for development and utilization of non-stationary signal analysis techniques, such as wavelets, or joint time-frequency analysis. Once sensor readings have been processed into a domain indicative of product performance, extraction of features most relevant to describing the product's performance can be accomplished in that domain. Thus, the method of feature extraction is essentially determined by the application and the domain into which sensor signals were processed.

Performance Assessment Module

The performance assessment module evaluates the overlap between the most recently observed signatures and those observed during normal product operation. This overlap is expressed through the so-called Confidence Value (CV), ranging between zero and one, with higher CVs signifying a high overlap, and hence performance closer to normal (Lee, 1995, 1996). In case data can be associated with specific failure modes, most recent performance signatures obtained through the signal processing and feature extraction module can be matched against signatures extracted from faulty behavior data. The areas of overlap between the most recent behavior and the nominal behavior, as well as the faulty behavior, are continuously transformed into CV over time for evaluating the deviation of the recent behavior from nominal to faulty.

Realization of the performance evaluation module depends on the character of the application and extracted performance signatures. If significant application expert knowledge exists, simple but rapid performance assessment based on the feature-level fused multi-sensor information can be made using the relative number of activated cells in the neural network, or by using the logistic regression approach. For products with open-control architecture, the match between the current and nominal control inputs and the performance criteria can also be utilized to assess the product's performance. For more sophisticated applications with intricate and complicated signals and performance signatures, statistical pattern recognition methods, or the feature map based approach can be employed.

Diagnostics Module

The diagnostics module tells not only the level of behaviour degradation (the extent to which the newly arrived signatures belong to the set of signatures describing normal system behaviour) but also how close the system behaviour is to any of the previously observed faults (overlap between signatures describing the most recent system behaviour with those characterizing each of the previously observed faults). This matching allows the Watchdog Agent to recognize and forecast a specific fault behaviour, once a high match with the failure associated signatures is assessed for the current process signatures, or forecasted based on the current and past product's performance. Figure 20.12 illustrates this signature matching process for performance evaluation.

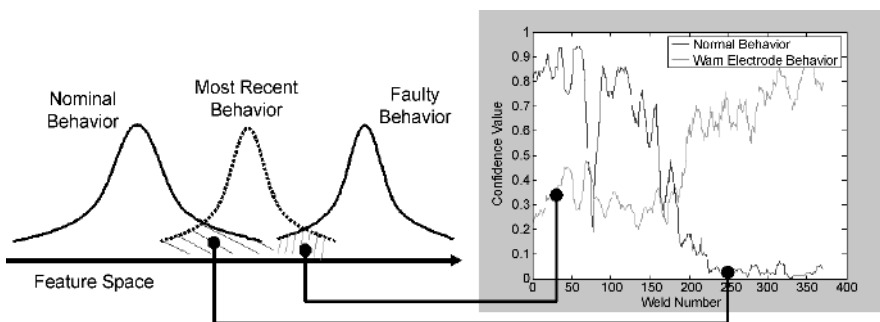


Figure 20.12. Performance evaluation using confidence value (CV) prediction and prognostics module

The prediction and prognostics module is aimed at extrapolating the behaviour of process signatures over time and predicting their behaviour in the future. Autoregressive moving average (ARMA) (Pandit and Wu, 1993) modelling and match matrix (Liu *et al.* 2004) methods are used to forecast the performance behaviour. Currently, ARMA modelling and Match Matrix methods are used to forecast the performance behaviour. Over time, as new failure modes occur, performance signatures related to each specific failure mode can be collected and used to teach the Watchdog Agent to recognize and diagnose those failure modes in the future. Thus, the Watchdog Agent is envisioned as an intelligent device that utilizes its experience and human supervisory inputs over time to build its own expandable and adjustable world model.

Performance assessment, prediction, and prognostics can be enhanced through feature-level or decision-level sensor fusion, as defined by the Joint Directors of Laboratories (JDL) standard of multi-sensor data fusion (Chapter 2, Hall and Llinas, 2000). Feature-level sensor fusion is accomplished through concatenation of features extracted from different sensors, and the joint consideration of the concatenated feature vector in the performance assessment and prediction modules. Decision-level sensor fusion is based on separately assessing and predicting process performance from individual sensor readings and then merging these individual sensor inferences into a multi-sensor assessment and prediction through some averaging technique.

20.5.4 Maintenance Decision Support System

In a complex industrial setup where maintenance planning still remains a difficult job, a system providing online support to a decision maker providing greater insight about the system would be of great value. This system could make recommendations to the decision maker but would not make the decisions. These kinds of systems are called Decision Support System. Work initially began on decision support systems (DSS) in the 1960s (Klein and Methlie, 1990). Research on DSS for industry-based maintenance of a single machine is found freely in the literature (Yam *et al.* 2001; Rao *et al.* 1990; Zhu, 1996; Tu, 1997; Fernandez *et al.* 2003; Yu *et al.* 2003). The main role of a DSS is to enhance decision making by an individual through easier access to problem recognition, problem structure, information management, statistical tools, and the application of knowledge (Santana, 1995). A number of computational tools are commonly used for DSS. Some of the tools are analytic hierarchy process (AHP) (Saaty, 1990; Davies, 1994; Bevilacqua and Braglia, 2000; Wang *et al.* 2007), knowledge based analysis (Liberatore and Stylianou, 1994), neural networks (Yam *et al.* 2001; Hurson *et al.* 1994), fuzzy logic, fuzzy networks (Schrunder *et al.* 1994; Mechefske and Wang, 2001), Bayesian theory (Keen, 1981; Charniak, 1991), and Petri nets (Jeng, 1997). Traditionally, decision support for maintenance was defined as a systematic way of selecting a set of diagnostic and/or prognostic tools to monitor the condition of a component or machine (Carnero, 2005). This type of decision support is necessary because different diagnostic and prognostic tools provide different ways to estimate and display health information. Therefore, users need a method for selecting the appropriate tool(s) for their monitoring purposes.

The Watchdog Agent toolbox integrates tools for equipment diagnostics and prognostics and provides basic methodologies for tool selections in order to help the maintenance manager make well informed decisions. The DSS commonly have three main functionalities: health assessment, condition diagnostic, and performance prediction. These are accomplished through several functional modules.

Planning maintenance for a production line is a complex task. Often models are developed based on statistical long-term behavior of the machines and maintenance actions are carried out in order to maximize the long-term benefits of the system. This approach shows good results in most cases, but is unable to capitalize on the additional opportunities which may emerge during regular operations of the plant. By considering both the immediate and the future rewards, the PMDSS (Figure 20.13) is developed to improve the system performance.

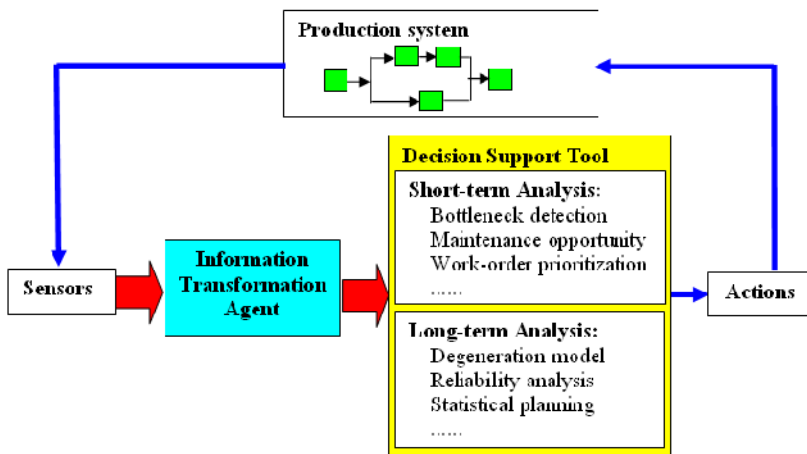


Figure 20.13. Framework for PMDSS

From a maintenance policy perspective, long term and short term are relative definitions. In general, it is difficult to define a period into short term or long term as it is dependent on the final objective, operating conditions, *etc.* For example, if failures occur frequently, a distribution or pattern may be used to describe the system's performance to study the long-term behavior. In contrast, if the failures are rare, then short-term analysis may turn out be more accurate and suitable than the statistical distributions.

A short term may be referred to as an operating period during which machines' failure behaviors cannot be assumed to be a statistical distribution, or the system cannot be analyzed as a steady state system; it could be hours, shifts, or days in a mass production environment. As seen in Figure 20.13, different tools are used for short-term analysis and long-term analysis. Short-term analysis depends heavily on the real time data and focus on the process control. The methods presently developed for such analysis include bottleneck detection, maintenance opportunity

planning, and maintenance task prioritization. On the other hand, the long-term study helps in the tactical level of decision making.

The working of PMDSS can be explained as follows: data is first received from various sensors installed on the production line. The data-information transformation system, such as Watchdog Agent, processes and transfers data into useful information. This information is further used to plan the maintenance and production. The long-term goal of the production lines is to meet the demand in a cost effective way. The long-term analysis ensures that the final goal is met, while the short term analysis ensures a continuous efficiency and also provides several opportunities for further improvement. Combining long term and short term analysis can lead to a smart final decision for improvement in system performance.

20.6 Technology Integration for Advanced E-maintenance

20.6.1 Generic ICT Interface

ICT advances in recent years have created a new landscape for implementing radically innovative solutions for maintenance and industrial asset management. The combined use of wireless sensors, networks, and mobile computing can fill in the information access gap that exists on the shop floor. Exploiting such technology advances requires greater effort to be devoted to integrating the information from various, distributed and heterogeneous sources. The key application scenario in this context is ubiquitous maintenance management (UMM). In UMM, maintenance-related information is seamlessly mediated back and forth throughout the different organization layers. It is made available at multiple locations (anywhere), instantaneously (anytime), to multiple users who are deemed to have authorized access to such information and to multiple operations and maintenance (O&M) subsystems, which operate on the basis of the current informational state of the organization and the shop floor (anyone).

Exploiting such opportunities enables a radical change in the landscape of maintenance services to occur. A maintenance service provider in the future will not be required to employ complex wired instrumentation and software situated in an isolated PC to gain access to machinery information and data acquired from the shop floor. Data relevant to condition monitoring and equipment maintenance can become ubiquitously available to technical personnel, *via* mobile and handheld devices, or even remotely *via* the internet. On the other hand, it becomes possible to reach operations decisions on the basis of timely, local and global production and assets state information.

The key application technologies, which permit this leap into an era with new e-maintenance solutions are advances in:

- Wireless networks;
- Sensor technology;
- Pervasive and contextualized computing; and
- Industrial information integration.

These advances are briefly discussed in the remainder of this section.

20.6.1.1 Wireless Networks

Wireless networks play a major role in the emerging e-maintenance and intelligent performance monitoring solutions. The deep penetration of wireless technology in modern industrial and consumer devices is based on the availability and growing maturity of wireless networking protocols. The family of common wireless protocols applicable to various forms of e-maintenance solutions include; *wireless PAN (WPAN) protocols* (mostly 802.15x) the established *WiFi (802.11x)*, the growing in pace *WiMax (802.16x)*, the emerging *MobileFi (802.20x)*, and the *post-3G protocols* related to mobile telephony. Such protocols are designed to serve different (but overlapping on some occasions) application needs, and each family of wireless protocols appears to have certain characteristics and strengths. They vary for instance in bandwidth, range of coverage, mobility support, quality of service, interference, and also costs.

WPAN wireless personal area networks are the wireless extension of Personal Area Networks. They are characterized by short range of coverage (typically from a few centimeters to a few meters), limited bandwidth, and energy consumption. They are mainly employed for establishing communication between peripheral devices (sensors, mobile computing devices, *etc.*), or for the exchange of information between the devices and a higher-level network. Among WPAN, one can distinguish several subcategories of the IEEE 802.15 protocols, namely the *IEEE 802.15.1 ("Bluetooth")*, *IEEE 802.15.3* (High data rate WPAN) and *IEEE 802.15.4* (Low data rate WPAN, ZigBee).

802.15.1 or Bluetooth provides connectivity between devices such as mobile phones, personal digital assistance (PDAs), laptops, PCs, *etc.*, over a secure but globally unlicensed and short-range radio frequency. Applications include control of and communication between a cell phone and a hands-free headset, wireless networking between PCs in a confined space and where little bandwidth is required, wireless communications between a PC and its peripherals, file transfer between devices, replacement of traditional wired serial communications in devices such as GPS receivers and control devices, replacement of IR communications, *etc.*

802.15.3 is targeting higher transfer rates PAN.

802.15.4 is targeting low cost and low rate WPAN. Due to its low power consumption and error resilience features, this protocol has gained great popularity in wireless sensor networks applications. Indeed, many vendors are now offering wireless sensing products, based on the ZigBee implementation of the 802.15.4 protocol.

WiFi One of the main advantages of *WiFi* family of *WLAN* protocols is the deep market penetration that already exists. The protocol is in fact supported by numerous vendors worldwide, and operates in the unlicensed spectrum of 2.4 GHz and 5 GHz bands. Higher data transfer rates of up to 600 Mbps will be supported by the latest and future generations of related protocols, such as 802.11n and 802.11s. Yet concerns exist with respect to its short range of coverage (in the order of tens of meters), and the lack of support for ensuring quality of service (QoS) in multimedia-rich transmissions, although QoS support has been added to the 802.11e implementation for supporting streaming applications.

WiMAX wide area coverage (in the order of kilometers) is a key feature of Wireless Metropolitan Area Networks of the *WiMAX* family. This family of protocols has far better QoS characteristics, compared to WiFi, allowing multimedia-rich transmissions with quality guarantees. Yet the theoretical transfer rates (in the order of several tens of Mbps) are still not considered likely to be achievable in the immediate future and in practice it is more realistic to anticipate rates not much higher than 10 Mbps. *WiMAX* is designed to operate both in licensed and unlicensed frequency bands, thus allowing potential interference and energy transmission restrictions to limit achievable performance. Nonetheless, predictions for *WiMAX* penetration are optimistic, indicating that overtaking WiFi is a probable scenario.

Wireless telephony beyond 3G developments towards mobile protocols beyond 3G, such as the evolution of UMTS (Universal Mobile Telecommunication System) into HSDPA (High-Speed Downlink Packet Access) and HSUPA (High-Speed Uplink Packet Access) or by combining those in HSPA (High-Speed Packet Access) are projected to achieve massive penetration by 2012. This allows data rates in the order of a few Mbps, ensuring fast streaming multimedia information, such as video clips (also see Krotov and Junglas, 2006).

A summary of the main characteristics of existing wireless networking standards applicable for e-maintenance or *intelligent maintenance solutions* are provided in Table 20.1 (adapted and expanded from Tan, 2006).

20.6.1.2 Sensor Networks

As mentioned earlier, a key factor in the successful implementation of e-maintenance or *Intelligent maintenance solutions* is the ability to perform Condition-Based Maintenance (CBM) efficiently. CBM requires that maintenance decisions are based on the identification of the current condition of monitored equipment. The implementation of efficient maintenance management strategies based on CBM presupposes that adequate condition monitoring, as well as machinery fault diagnostics and prognostics are in place. Current advances in sensor technology offer a growing range of choices for the use of wireless sensors. Such sensors are easier to deploy compared to their wired counterparts and facilitate the ubiquitous availability of sensorial data at the shop floor. Wireless sensor data networking is typically performed via 802.15.4 enabled sensors and devices. Several vendors already offer wireless sensor solutions (e.g., Crossbow sensors, Mica and MicaZ motes). Application development is based, for example, on Berkeley's TinyOS and NesC. Greater interoperability is likely to be offered by the newly introduced platform by SUN, namely SPOT (Small Programmable Object Technology) and the SQUAWK Java Virtual Machine, designed for interoperability and operation on embedded system devices. SPOT technology integrates 802.15.4 wireless connectivity and is offered as part of the Java Micro Edition platform (J2ME). Coupling sensor technology with RFID technology offer additional capabilities for performing rapid asset, equipment and component tracking and linking sensorial information to the data collection point, i.e., directly providing the appropriate context for the collected information. Such developments enable the seamless integration of networked and embedded sensing and

computing devices, making information from the shop floor machinery ubiquitously available to the networked enterprise.

Table 20.1. Characteristics of trends in mobile technologies (adapted and expanded from Tan and Wond, 2006)

	3G Derivatives	WPAN	WiFi	WiMAX
Standard	HSDPA/HSUPA, EV-DO	802.15	802.11x	802.16x
Features				
Maximum bandwidth (lower effective bandwidth in practice)	14.4 Mbps (HSDPA), 5.8 Mbps (HSUPA), 46.5 Mbps (EV-DO Rev. B using all carriers)	Up to 55 Mbps for 802.15.3 with a goal of exceeding 110Mbps (10 m range) or 400 Mbps (5 m range). Up to 2.1Mbps for Bluetooth 2.0. Up to 250Kbps for 802.15.4.	11 Mbps (b), 54 Mbps (g), over 100Mbps (n)	Over 70 Mbps (fixed), 15 Mbps (mobile)
Operations	Cellular operators	Personal area networks, cellular phone peripherals, wireless sensor networks.	Individuals, Wireless Internet Service Providers (WISPs)	Individuals, Wireless Internet Service Providers (WISPs), wireless operators
License	Yes	No	No	No/yes (optional)
Range	Several km	Typically up to 10 m	100m maximum in best conditions	50 km in best conditions
Pros	Range, mobility	Low power consumption, QoS, low cost	Bandwidth, costs	Bandwidth, range, mobility
Cons	Price	Lower bandwidth	Short range, poor quality of service (QoS)	Interference issues in unlicensed bands

20.6.1.3 Mobile Computing and Context Awareness

The integration of wireless technologies with sensor and mobile computing devices enable the application of mobile and contextualized computing concepts to serve modern maintenance engineering practice. The most typical application usage is the ubiquitous information mediation, which may be contextualized too, *i.e.*, relevant to specific user profiles, locations, activities, or assets. Mobile technologies can also offer location based services (LBS). LBS can be of great

benefit for a wide range of applications, including indoor or outdoor navigation aid and location-contextualized content delivery. Other typical applications are in logistics, where asset tracking and handling of information relevant to the asset can become automated, while at the same time offered services can be made adaptive, to fit better to different user profiles.

Contextualized or situated computing offers innovative ways of mediating information and providing adaptive user interfaces, based on the apparent context of each service request. A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task. Typically, context may be determined by location-specific information or data, time, user profiles and identity, as well as by the characteristics of the specific activity performed. Let's consider a typical ubiquitous maintenance scenario, wherein a production engineer is in front of specific production machinery. The engineer carries a handheld device (PDA), equipped with a RFID tag reader. The monitored machinery is itself equipped with a RFID tag and several sensors to monitor its operating condition. The plant back office stores structured and historical information related to the production plans and constraints, as well as to the overall plant and machinery operation. The RFID enabled PDA identifies the monitored machinery and is therefore able to associate measured information with either locally (on the PDA) stored data or compare against historical data which is centrally stored. Furthermore, once the monitoring context (machinery, production type, *etc.*) has been identified, the engineer on the shop floor may gain direct access to technical documentation relevant to the monitored machinery, thus facilitating the assessment of the monitoring situation. Beyond this, the monitored data may be processed by a condition monitoring, diagnostics and prognostics system, thus directly offering expert advice on site. Depending on the level of the information integration, this advice may result from processing the collected data in the light of current and future production and maintenance constraints. In other words, maintenance actions are guided by context-aware decision and mobile computing support. Clearly, reaching informed maintenance decisions and prompting personnel to specific maintenance actions can greatly benefit from the capability to operate within a context-aware or situated computing decision support environment. The very notion of context is in itself quite broad and may influence the operation of networked devices. A recent example is context-aware sensing devices, wherein sleep or awake sensor operation and signal transmission is determined by the application context, thus enabling energy saving operation. Arguably, maintenance engineering practice and e-maintenance applications have much to gain from employing similar context-aware mobile computing concepts. With information integration, e-maintenance can only exploit current technology in wireless networking, sensing and mobile computing insofar as the information provided by the various heterogeneous sources is appropriately integrated. This technology barrier should not be underestimated, as numerous past attempts to advance maintenance engineering practice have had limited impact, since there was no provision for data interoperability and information integration. The applicable standards in the machinery operations and maintenance domain are to some extent being integrated under the auspices of the MIMOSA association in collaboration with the OPC Foundation and the ISA-SP95 committee. This is also

steering the development of the related ISO standard 'ISA S95 Standard for Enterprise-Control System Integration' whose first section is the ISO/IEC 62264 standard. The above three organizations are collaborating in developing the open O&M standard which is effectively a standard integrating and complementing a group of associated standards. In particular a common information bus is defined and two primary sets of interfaces are also defined to operations on one side and to enterprise systems on the other. Pertinent standards for interfacing on the operations side include:

- OPCXML and MIMOSA OSA-EAI for low level accessing of machine control systems and data;
- ISA SP95, OPC XML and OSA-EAI for intermediate, plant level forecasting, planning and scheduling systems;
- ISA SP95 for materials and personnel management at the logistics level; and
- OSA-EAI for interfacing to all types of physical asset resource management systems.

A complementary set of standards are the W3C managed XML and Semantic web standards that supply a basis for interoperable semantic information exchange and ontological representations. The adoption and implementation of such standards is likely to lead to enhance technologies and information integration prospects, which may ultimately lead to a successful e-maintenance implementation strategy.

20.6.2 Generic Interface Requirements for Watchdog Agents

The open architecture Watchdog Agent toolbox uses the development procedure shown in Figure 20.14. The interface solutions for a comprehensive Watchdog technology, in general, comprise hardware applications, software applications, and other user interface solutions as discussed below.

20.6.2.1 Hardware

For a certain industry application, the selection of Watchdog Agent hardware depends on characteristics of the input/output signals (*e.g.*, what type of input/output signal and how many channels needed), which tools or algorithms are selected (*e.g.*, different algorithms require different hardware computation and storage capacities), and the hardware's working environment (for example, which decides the hardware's storage type, temperature range, *etc.*). The hardware prototype currently used in the IMS Center is based on PC104 architecture, as shown in Figure 20.15a. PC104 architecture enables the hardware to be easily expanded to a multi-board system, which includes multiple CPUs and a large number of input channels. It has a powerful VIA Eden 400MHz CPU and 128MB of memory since all of the tools are embedded into the hardware. It has 16 high speed analog input channels to deal with highly dynamic signals. It also has various peripherals that can acquire non-analog sensor signals such as RS-232/485/432, parallel and USB. The prototype uses a compact flash card for storage, so it can be placed on top of machine tools and is suitable for withstanding

vibrations in a working environment. Once a certain set of tools/algorithms is determined for a certain industry application, commercially available hardware, such as Advantech and National Instruments (NI) as illustrated in Figure 20.15 b, c respectively, will be further evaluated for customized Watchdog Agent applications.

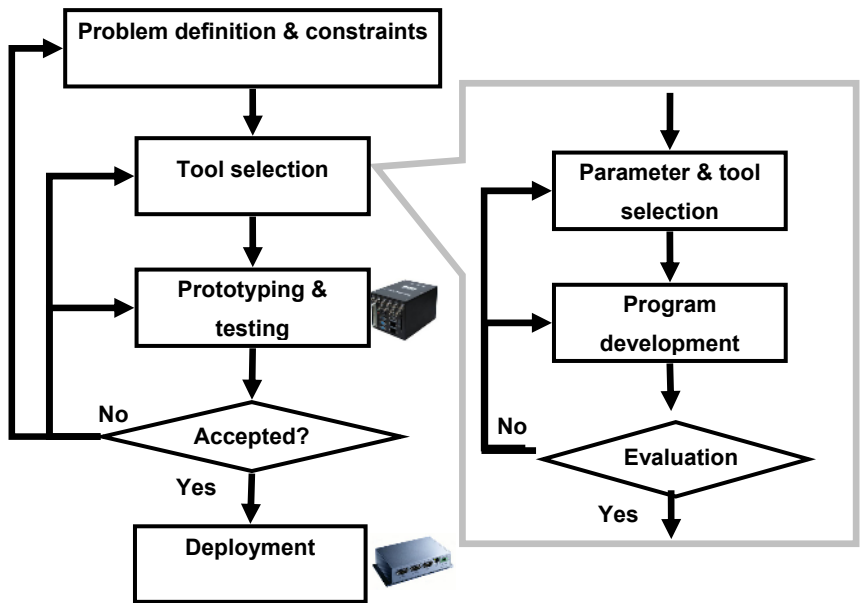


Figure 20.14. Flowchart for developing Watchdog Agent tools



(a) IMS prototype hardware (b) Advantech UNO-2160 (c) NI-CompactRIO

Figure 20.15. Options of hardware prototypes for Watchdog Agent application

20.6.2.2 Software Development

The software system of the Watchdog Agent-based IMS platform consists of two parts: the embedded side software and the remote side software, as shown in Figure 20.16. The *embedded side software* is the software running on the Watchdog Agent hardware, which includes a communication module, a command analysis module, a task module, an algorithm module, a function module, and a DAQ module. The communication module is responsible for communicating with the remote side *via* TCP/IP protocol. The command analysis module is used to analyze different commands coming from the remote side. The task module includes multi-thread scheduling and management. The algorithm module contains specific Watchdog

Agent tools. The function module has several auxiliary functions such as channel configuration, security configuration, and email list and so on. The DAQ module performs A/D conversion using either interrupt or software trigger to get data from different sensors. The *remote side software* is the software running on the remote computers. It is implemented by ActiveX control technology and can be used as a component of the Internet Explorer browser. The remote side software is mainly composed of a communication module and a user interface module. The communication module is used for communicating with the embedded site *via* TCP/IP protocol. The user interface has a health information display, an ATC status display, and a discrete event display. It also possesses an algorithm module, as well as error log database and data format interface.

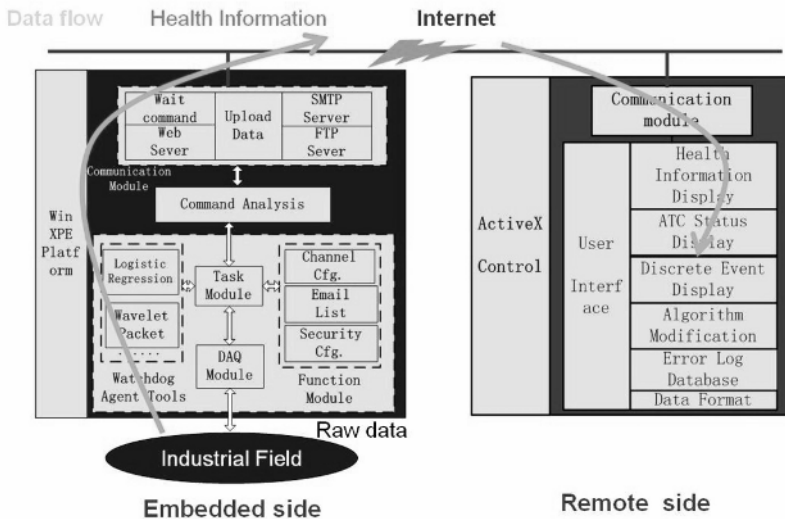


Figure 20.16. Software structure of Watchdog Agent

20.6.2.3 Remote Monitoring Architecture and Human Machine Interface Standards

A generic four-layer infrastructure for remote monitoring and human machine interface standards is illustrated in Figure 20.17. The *data acquisition layer* consists of multiple sensors, which obtain raw data from the components of a machine or machines in different locations. The *network layer* will use either traditional Ethernet connections, or wireless connections for communication between the Watchdog Agents, or for sending short messages (SM) to an engineer's mobile phone *via* GPRS services. The *application layer* functions as a control server to save related information and control the behavior of the Watchdog Agents in the network. The *enterprise layer* offers a user-friendly interface for maintenance-related engineers to access information either *via* an internet browser or a mobile phone.

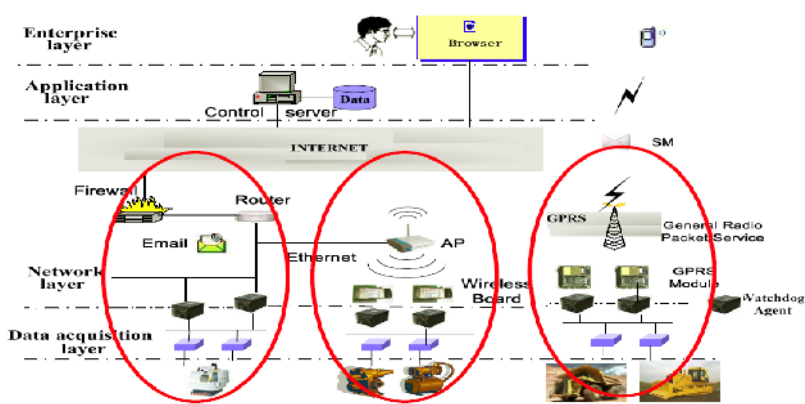


Figure 20.17. Illustration of Watchdog Agent-based remote monitoring architecture

20.6.3 Systems-user Interface Needs

As the use of advanced technologies for e-maintenance solutions gradually progresses, there is also an emerging critical requirement for better systems-user interfaces. Such an interface, in the first place, attempts to harmonize the interactivity and communication between the *technological platform* and the *social setup* or the *user environment* within which the technology is operationally embedded. This has important implications on the reliability, safety, and even security of the application environment. Types of interfaces and their features are shown in Table 20.2.

Years of R&D activities in different contexts have resulted in the proposition and application of numerous techniques to enhance systems-user interface performance. Some of the popular interface applications, that are equally applicable to *Integrated* e-maintenance solutions, include the following (see also Eason, 1988; Bannon, 1990; Wickens, 1992; Bailey, 1996; Booher, 2003; Clarke *et al.* 2003; Gunasekaran *et al.* 2003; Zimmermann *et al.* 2005; Wikipedia, 2007);

20.7 Some Industrial Applications

20.7.1 E-maintenance Solutions for Complex Industrial Assets

As the complexity of industrial asset management process grows, the owners or operators of production, manufacturing, and process plants and facilities have begun to seek innovative technical solutions to reduce asset related risks. The oil and gas (O&G) exploration and production industry on the *Norwegian Continental Shelf (NCS)* today provides a specific example in this context. The entire industry is on the way to establish what is called '*Smart assets*'. Within this dedicated initiative some major initial steps are also seen to have been taken to implement and use comprehensive e-maintenance solutions for offshore assets.

Table 20.2. Relevant systems-user interface issues for complex e-maintenance solutions

Type of interface	Interface feature
Textual user interface (TUI)	The interface is built using appropriate texts and symbols
Command-line interface (CLI)	This is to enable more effective users interaction with their operating systems by means of a command line interpreter. The command line interpreter reads the textual inputs given by the user and interprets them
Graphical user interface (GUI)	This interface involves various graphical icons, visual indicators or widgets, mainly allowing user interaction with computers and/or computer-controlled devices. This can be seen combined with text-based navigation aids for the user
Interaction design (IxD)	This seeks to understand and specify the user needs and then to bring those user-need specifications to the design of systems. It is the way to enhance the usability of technological solutions through better integration of experiences
Experiential design	Here the focus is mainly on the more active utilization of human experience, that involves customs, skills, knowledge, expertise, beliefs, perceptions, wants, <i>etc.</i> in the design of complex systems and environments
Information design	This has the main intention of making the information available and present them to the user in a way that he/she can make decisions and perform assigned tasks in a more effective and an efficient manner
Graphic design	This seeks to actively use various types of graphics and images for the purpose of enhancing visual communication and information exchange between the user and a technical system
User centred design (UCD)	This by far is a methodology where a comprehensive analysis of actual end-users characteristics, requirements and limitations is done in order to enhance the user-interface. The knowledge and understanding about the user is taken into each step of the design process

Two major issues in particular have contributed much to the design and development of *Smart assets* and to explore the feasibility of advanced e-maintenance solutions, namely:

- Gradually declining oil production, and the finding of more marginal fields; and
- Rising operating costs, particularly due to ageing equipment, systems, and more technically challenging development projects.

In this setting, the risk exposure became more and more obvious a few years ago. It triggered major political and business sources to explore various feasible options to reduce commercial risks. Subsequently, the most favoured solution, termed *Integrated eOperations*, was adopted in early 2000s, which so far has resulted in USD billions of investments.

Today, the *Integrated e-Operations* concept is in a fast track of progress through an industry wide re-engineering process. This mainly incorporates:

- High-tech embedded control/support/coordinating centres to enhance connectivity and interactivity between offshore O&G assets and Onshore support systems;
- Advanced fiber-optic based and wireless ICT systems, and web-based solutions for real-time data traffic and communication;
- Standardized technical platforms, for instance based on semantics and ontologies, for data exchange between business partners;
- Novel data acquisition and interpretation technologies, 3-D technologies, and simulation solutions for rapid decision support; and
- Novel ICT solutions to support online collaborative decisions and activity planning between business partners (*e.g.*, operators, engineering contractors, drilling service providers, CBM experts, *etc.*) to improve work processes.

A unique feature of ongoing developments is that it seeks for integrated solutions for reliable, safe and secure *24/7 online real-time* operations in offshore assets.

Maintenance management has drawn major attention in this setting particularly owing to its impact on the operating costs, production availability, health & safety, and environmental performance. Major benefits are mainly expected particularly through the use of e-maintenance solutions where CBM applications remain the bedrock. Traditionally, CBM has not been that attractive for offshore applications, but with the current advancement in application technologies, particularly the ICT sector, it appears that a new avenue of growth has been opened up. The integral components that have set up the necessary technical foundation for advanced e-maintenance solutions on NCS includes (see also Figure 20.18):

- Rapid application in sensor technologies for data acquisition;
- Implementation and accelerated developments in the large ICT network termed '*Secure Oil Information Link*' (*SOIL*) based on fiber-optic net for real-time data exchange;
- High-tech CBM expert centres onshore for data analysis, interpretation, and troubleshooting;
- Enhanced use of smart video-based online communication technologies such as Visi-Wear for online interaction between offshore asset crew, onshore support engineers, and remotely located CBM experts; and
- wireless applications and Web-based solutions for establishment of supportive expert networks.

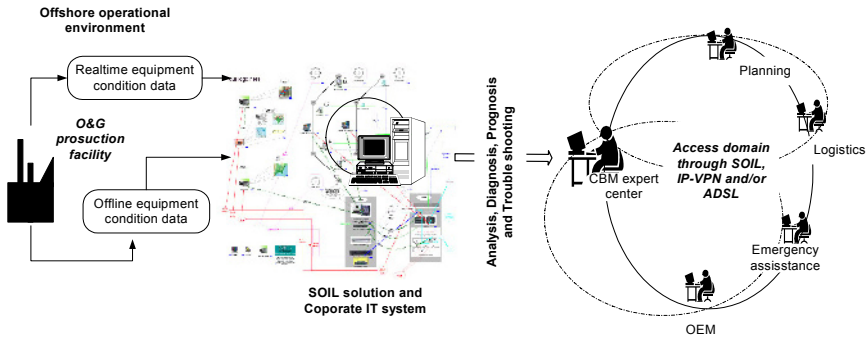


Figure 20.18. Emerging environment and operational landscape for e-aintenance solutions on NCS

The technological impact in this context has been quite substantial as discussed in Liyanage and Langeland (2007) and Liyanage *et al.* (2006). While the technology has shown its own pace of development and potential to realize 24/7 online real-time operations, the extensive re-engineering process on NCS and the considerable changes in working patterns and organizational forms have now drawn the attention on three specific aspects as shown in Figure 20.19.

There are some ongoing R&D efforts at the moment to address a number of aspects related to integration, interfacing, and coordination issues. Yet presumably, more comprehensive fail-proof solutions cannot be expected before 2010 or so.

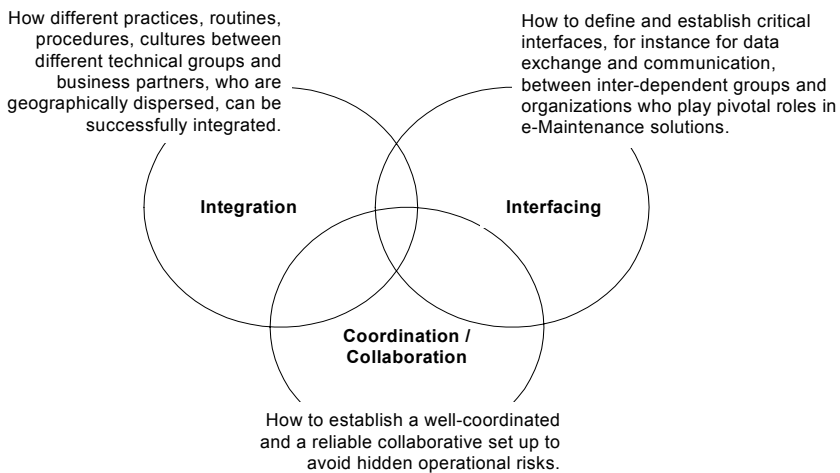


Figure 20.19. Important aspects beyond technology for robust e-maintenance solutions

20.7.2 Watchdog Technology for Product Life-cycle Design and Management

In the area of Collaborative Product Life Cycle Design and Management, the Watchdog Agent can serve as an infotronics agent to store product usage and end-of-life (EOL) service data and to send feedback to designers and life cycle management systems. Currently, an international intelligent manufacturing systems consortium on product embedded information systems for service and EOL has been proposed. The goal is to integrate Watchdog Agent capabilities into products and systems for closed-loop design and life cycle management, as illustrated in Figure 20.20.

The R&D activities will continue advancing current research to develop technologies and tools for closed-loop life cycle design for product reliability and serviceability, as well as explore research in new frontier areas such as embedded and networked agents for self-maintenance and self-healing, and self-recovery of products and systems. These new frontier efforts will lead to a fundamental understanding of reconfigurability and allow the closed-loop design of autonomously reconfigurable engineered systems that integrate physical, information, and knowledge domains. These autonomously reconfigurable engineered systems will be able to sense, perform self-prognosis, self-diagnose, and reconfigure the system to function uninterrupted when subject to unplanned failure events.

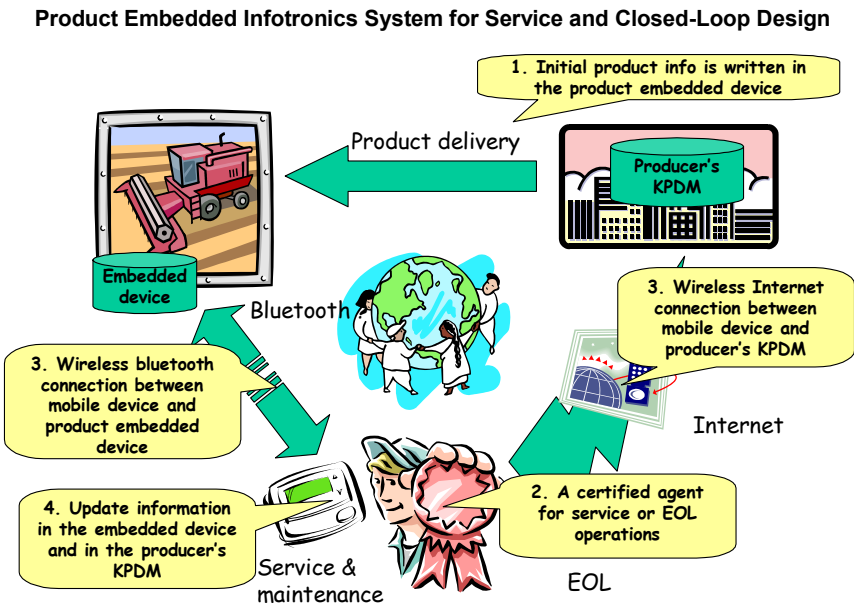


Figure. 20.20. Embedded and tether-free product life cycle monitoring

20.7.3 Watchdog Technology to Trouble-shoot Bearing Degradation

20.7.3.1 Signal Processing and Feature Extraction

For sustained defects, Fourier-based analysis, which uses sinusoidal functions as basis functions, provides an ideal candidate for extraction of narrow-band signals. When dealing with a discrete, or a sampled/digitized analogue signal, Discrete Fourier Transform (DFT) is the appropriate Fourier analysis tool. DFT can be computed efficiently in practice using a fast Fourier transform (FFT) algorithm. In this chapter, FFT is presented to extract features from the vibration signals. Energy of the sub-band around each bearing defect frequency in the frequency spectrum is calculated as the feature for the health assessment algorithm.

By using the FFT algorithm, the vibration signal is translated from time-domain into its equivalent frequency domain representation. The magnitude spectrum can be subdivided into a specific number of sub-bands. A sub-band is basically a group of adjacent frequencies. The center frequencies of these sub-bands have already been pre-defined at the bearing defect frequencies such as ball passing frequency inner-race (BPFI), ball passing frequency outer-race (BPFO), ball spin frequency (BSF), and foundation train frequency (FTF). The energy in each of these sub-bands is computed and passed on to the health assessment algorithm in next step.

The feature vector that describes the health state of a bearing is (energy around the bearing defect frequencies, amplitude peak at 1X RPM, amplitude peak at 2X RPM, amplitude peak at 3X RPM, acceleration maximum, RMS, kurtosis).

20.7.3.2 Fault Diagnosis

With available data from different bearing failure modes, the method of self-organizing map is applied to provide a health map in which different regions indicate different defects of a bearing.

In this industrial case, vibration data was collected from an experiment in which three bearings are artificially made to have a roller defect, an inner-race defect, and an outer-race defect respectively. Vibration data is also obtained from a bearing with no defect. Figure 20.21 shows the vibration signals of the four types of spindle bearing states.

After training, a health map for the classification of different bearing failure modes is obtained, as shown in Figure 20.22. The health map shows four regions which are labeled 'N', 'RF', 'IF', and 'OF' that indicate the normal status, roller defect, inner-race defect, and outer-race defect respectively. During the degradation process a bearing may stochastically develop different failure patterns depending on the physical structure and the operating condition of the machine. The health map shown in Figure 20.22 can also be used to detect the degradation process if the bearing develops different failure patterns. When the bearing is in normal condition, the hit points of the testing data will be located near the 'N' area on the health map. If it develops an outer-race defect, the hit points of the testing data will migrate to, 'OF' area on the map.

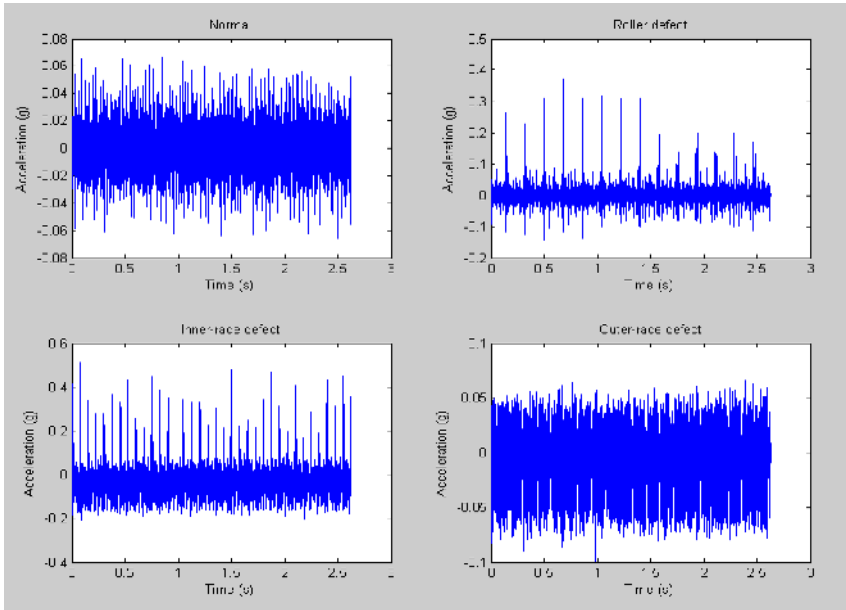


Figure 20.21. Vibration signals of the four types of spindle bearing states

20.7.3.3 Health Assessment

The SOM method is applied to the extracted feature vectors to draw conclusions concerning bearing degradation condition.

Typically, only measurement at normal operating conditions is available. In rare cases, there exist historical data of the development of defects in measurements of a complete set of all possible defects. SOM can be used to evaluate the bearing health condition when only normal data is available. After a description of normal machine behavior is set up, anomalies are expected to appear as significant deviations from this description.

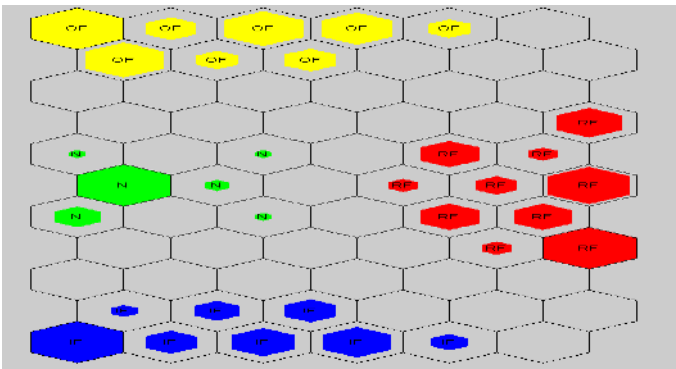


Figure 20.22. Health map for classification of different bearing failure patterns (N: normal condition, OF: outer-race failure, RF: roller failure, IF: inner-race failure)

SOM can be used to identify the different health states of a bearing during its whole life cycle. A whole life cycle of a bearing can be generally considered as normal stage, initial defect and fault propagation stage (or critical stage) and faulty stage. A run-to-failure test is carried out for the bearing test. The test is stopped when a significant amount of metal debris is found and the bearing develops a roller defect. Features are extracted from the vibration data using FFT, and the features are used as input vectors to train the SOM. After the training, two maps can be obtained and are shown in Figure 20.22.

The map on the left is the U-matrix (Unified distance matrix) map for the entire baseline training dataset. U-matrix map visualizes distances between neighboring map units, and aides in perceiving the cluster structure of the map: areas of high values (shown by dark color) of the U-matrix indicate a cluster border; areas of low values (shown by light color) indicate clusters themselves. In this map, three different areas are separated by the darker hexagons (boundary in the map), which indicates a right classification of the three training datasets. In the map on the right, three areas in different locations are indicated by different colors. Those areas represent the normal, critical and faulty stages of the bearing respectively. This map can be used as a health map to test the new data set. The test result is indicated by the 'hit point' on the map. The area in which the 'hit point' is located represents the condition stage of the bearing. In Figure 20.23, the 'hit point' located in the normal area indicates that the bearing is in normal stage.

For each input feature vector, a BMU can be found in the SOM. The distance between the input data feature vector and the weight vector of the BMU, which can be defined as the minimum quantization error (MQE), actually indicates how far away the input data feature vector deviates from the normal operation state. Hence, the degradation trend can be visualized by the trend of the MQE. As the MQE increases, the extent of the degradation becomes more severe.

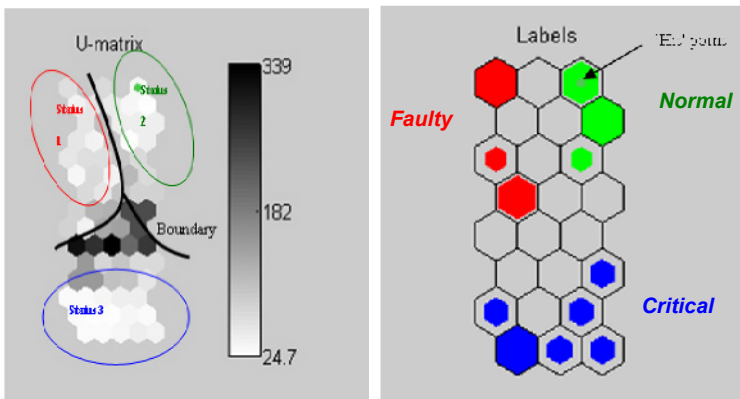


Figure 20.23. Health map of different stages of the bearing with roller defect

Data from the first 1000 cycles during which the bearing is in normal condition are used to train the SOM. After the training, the whole life cycle data of the bearing with roller element defect is used for testing and the corresponding MQE

values are calculated. From the curve shown in Figure 20.24, the degradation process of the bearing can be clearly observed. In the first 1000 cycles, the bearing health is in good condition and the MQEs are near zero. From cycle 1250 to cycle 1500, the initial defects appear and the MQE begins to increase. The MQE keeps increasing until it nears cycle 1700, which means the defects become more serious. Subsequently, until approximately cycle 2000, the MQE drops because the propagation of the roller counterbalances the vibration. The MQE will increase sharply after this stage till the bearing fails.

20.8 Challenges of E-maintenance Application Solutions

Smart technological solutions, as adopted by today’s industries, are key to mitigating new sets of operational risks and to realizing commercial benefits. Over the last few years, the maintenance discipline has been inundated with numerous forms of such solutions. Ahead of these developments, obviously, remain the CBM applications. The subject matter, as of today, has grown to a considerable extent, suggesting interesting troubleshooting methods for instance including neural networks, expert systems, fuzzy logic, genetic algorithms, multi-agent platforms and case based reasoning, facilitating the path towards advanced *intelligent* e-maintenance solutions.

In the e-maintenance context, Watchdog Agents and Intelligent Maintenance Solutions have recently drawn major attention not only due to their technological marvel but also owing to their specific potential to provide more comprehensive solutions. Despite some developments, more robust solutions for industrial applications (as in Figure 20.25) are yet to be introduced.

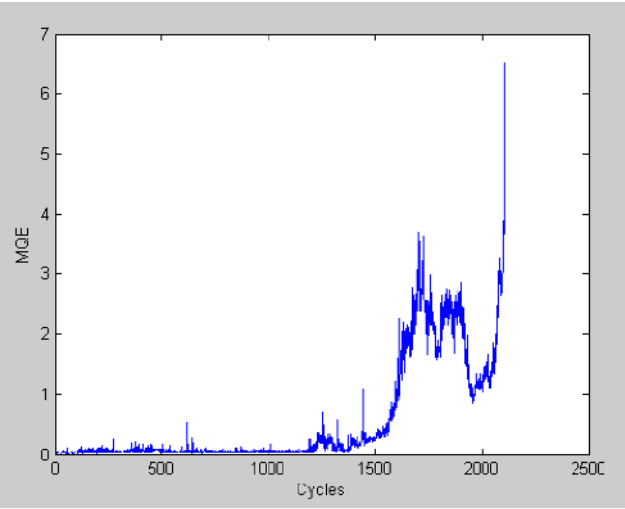


Figure 20.24. MQE of the degradation process of the bearing with roller defect

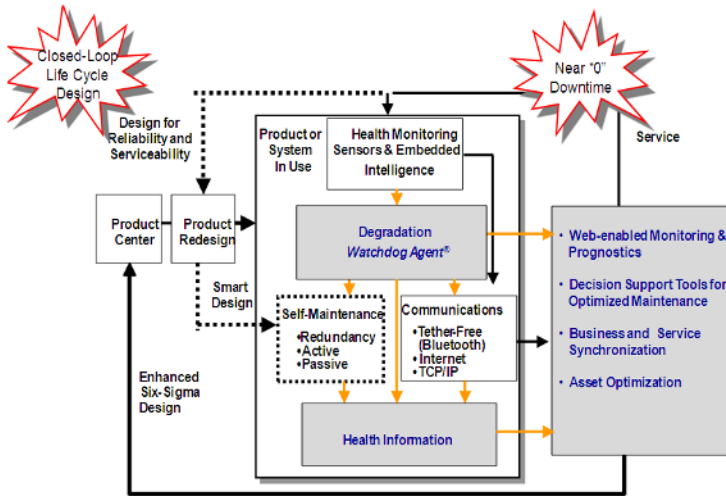


Figure 20.25. Key elements of a robust intelligent maintenance system

Some of the specific challenges of comprehensive e-maintenance application solutions relate to;

- Advanced maintenance simulation software for maintenance schedule planning and service logistics cost optimization for transparent decision making;
- Integration of decision support tool and optimization techniques for proactive maintenance based on sustainable and self-aware artificially intelligent system that learns from its own operation and experience; and
- Wireless sensor network made of self-powered, or very low energy consumption wireless motes for machine health monitoring and embedded prognostics.

Obviously these technologies are very critical for monitoring equipment or systems in a complex environment where availability is the major concern.

In addition, there exist issues related to the impact of new technologies for innovative e-maintenance solutions. For instance, major concerns can be related to reliability verifications, safety exposure, and security of advanced e-maintenance solutions that are under implementation for commercial advantages, and also thus the potential risk exposure induced by them towards the operators/asset owners. The risk exposure as such is enhanced due to the extensive expansion of formal organizational forms relying more and more on business alliances and networks. Some of the issues in this context, for instance related to organizational factors in CBM applications (Bengtsson, 2004) and wider socio-technical systems level implications (Liyanage, 2006, Liyanage and Bjerkebak, 2007) have already been highlighted. The principal concern has been that when the complexity of technological solutions increases, such as e-maintenance, a more professional approach is necessary to ensure that hidden threats and vulnerabilities are

identified. False alarms, interpretation errors, human psychological capacities, *etc.*, have major roles in this context to ensure reliable, fail-safe and error-free operations in industrial facilities. This still remains a hidden challenge and is mostly left inadequately addressed.

Obviously, an e-maintenance vision does not depend solely on the new technological capabilities. E-maintenance also demands innovative solutions with respect to the establishment of a robust and a reliable service infrastructure. This underlines the need for open platforms to help support technology and service integration. Some applications in this regard are currently visible in different industrial sectors, for instance oil and gas production, equipment design and manufacturing, *etc.* Some early work in this context includes Yu *et al.* 2003; Lung and Marquez 2006, and Muller *et al.* 2008.

In principal the term ‘e-maintenance’ not only implies a complex assembly of advanced technologies. A ‘live’ system is required to support the productive use of application technologies. Obviously, a system is more than the addition of all the components. In this context, a robust e-maintenance system require establishment of an effective interface between four critical elements, namely;

- Semantics and ontologies for effective data management;
- Organizational solutions;
- Technological applications; and
- Work process redefinition.

This is to ensure interoperability is achieved through seamless integration of various dynamic components, beginning from the intelligent solution specifications to the implementation and exploitation process to manage better the equipment and plant condition.

Notably, application solutions pertaining to e-maintenance are still on the verge of growth. However, in the emerging cost competitive and time critical industrial environments, e-maintenance appears to be the way forward that allows a given industrial environment. This helps in exploiting the technological advancement to better manage risks and vulnerabilities through effective knowledge and information management, as well as collaborative problem solving and learning processes.

20.9 Conclusion

Intelligent maintenance solutions and e-maintenance applications have drawn much attention lately both in academia and industry. This is perhaps attributable to the need for innovative and cost-effective solutions when industry is faced with novel challenges. Obviously, the development on the subject matter so far, both in research and development as well as application fronts, has brought much hope. Yet the challenges are also quite numerous owing to the complexity of the underlying issues, and the principal needs for systems solutions based on effective interfaces.

References

- Akyildiz IF, *et al.* (2002). Wireless Sensor Networks: A Survey. *Comput Netw J* 38(4): p. 393–422.
- Badia FG, Berrade MD and Campos CA (2002) Optimal Inspection and Preventive Maintenance of Units with Revealed and Unrevealed Failures. *Reliab Eng and Syst Saf* 78: 157–164.
- Bailey RW (1996) Human performance engineering, Prentice Hall.
- Bannon LJ (1990) A pilgrim's process: From cognitive science to cooperative design, AI and Society
- Barbera F, Schneider H and Kelle P (1996) A Condition Based Maintenance Model with Exponential Failures and Fixed Inspection Interval. *J of the Oper Res Soc* 47(8): 1037–1045.
- Bengtsson M (2004) Condition Based Maintenance Systems – An investigation of technical constituents and organizational aspects. Licentiate thesis, Mälardalen University, Eskilstuna, Sweden.
- Bevilacqua M and Braglia M (2000) The analytical hierarchy process applied to maintenance strategy selection. *Reliab Eng and Syst Saf* 70: 71–83.
- Bonissone G (1995) Soft computing applications in equipment maintenance and service, ISIE '95, Proceedings of the IEEE International Symposium, 2: 10–14.
- Booher HR (2003) Handbook of human systems integration, Wiley-Interscience.
- Bose S and Sohi S. (2006) Enabling Seamless Contextual Collaborations for Mobile Enterprises. in 4th International Workshop on Distributed and Mobile Collaboration (DMC 2006), University of Manchester, UK.
- Brotherton T, Johnson T, Chadderdon G (1999) *Classification and Novelty Detection Using Linear Models and a Class Dependent-Elliptical Basis Function Neural Network*. Proceeding of: Int'l Conf on Neural Nets, Anchorage, AK, May 1999.
- Bruns P (2002) Optimal Maintenance Strategies for Systems with Partial Repair Options and without Assuming Bounded Costs. *Eur J of Oper Res* 139: 146–165.
- Bunday BD (1991) Statistical Methods in Reliability Theory and Practice, Ellis Horwood.
- Burrus C, Gopinath R, Haitao G (1998) Introduction to wavelets and wavelet transforms - a primer. NJ: Prentice Hall.
- Campbell DA, Sarker S, Valacich JS, (2006) Collaboration using mobile technologies: When is it essential? in Proceedings of the International Conference on Mobile Business (ICMB'06). Copenhagen, Denmark: IEEE.
- Carnero M C (2005) Selection of diagnostic techniques and instrumentation in a predictive maintenance program: A case study. *Decision Support Systems*, 2005, pp. 539–555.
- Casoetto N, Djurdjanovic D, Mayor R, Lee J, Ni J (2003) Multisensor process performance assessment through the use of autoregressive modelling and feature maps. *Trans of SME/NAMRI* 31: 483–490.
- Cavory G, Dupas R, and Goncalves R (2001) A Genetic Approach to the Scheduling of Preventive Maintenance Tasks on a Single Product Manufacturing Production Line. *Int J of Prod Econ* 74: 135–146.
- Cellular News, (2006) One billion high speed data users by 2012, 15 February 2006.
- Charles AS, Floru IR, Azzaro-Pantel C, Pibouleau L and Domenech S (2003). Optimization of Preventive Maintenance Strategies in a Multipurpose Batch Plant: Application to Semiconductor Manufacturing. *Comput and Chem Eng* 27: 449–467.
- Charniak E (1991) Bayesian networks without tears. *AI Magazine*, 12, pp. 50–64.
- Chen D and Trivedi K (2002) Closed-Form Analytical Results for Condition-Based Maintenance. *Reliab Eng and Syst Saf* 76: 43–51.
- Chen CT, Chen YW and Yuan J (2003) On a Dynamic Preventive Maintenance Policy for a System under Inspection. *Reliab Eng and Syst Saf* 80: 41–47.

- Chen N, *et al.* (2005). A Framework for Ambient Applications — Context-Based Autonomous Extension of Applications. in 2005 Systems Communications, Montreal, Canada: IEEE.
- Chong SK, Loke SW, Krishnaswamy S (2005) Wireless Sensor Networks: From Data to Context to Energy Saving. 33–42. in 2005 International Workshop on Ubiquitous Data Management (UDM 2005). IEEE Computer Society 2005.
- Clarke S, Coakes E, Hunter MG, *et al.* (2003) Socio-technical and human cognition elements of information systems, Information Science Publishing.
- Cohen L (1995) Time-frequency analysis. NJ: Prentice Hall.
- Coley, D. A., (1999). An Introduction to Genetic Algorithms for Scientists and Engineers World Scientific Co. Pte. Ltd
- Cox D. (1972). Regression models and life tables (with discussion). J of the Royal Stat Soc ser B 34: 187–220.
- Davies MAP (1994) A multicriteria decision model application for managing group decisions. J of the Oper Res Soc 45 (1), pp 47–58.
- Dekker R (1995) Integrating optimization, priority setting, planning and combining of maintenance activities. Eur J of Oper Res 82: 225–240.
- Dey AK, Abowd GD (1999) Towards a Better Understanding of Context and Context-Awareness. in Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing 1999. Karlsruhe, Germany: Springer-Verlag London, UK.
- Djurdjanovic D, *et al.* (2000) Computerized classification of temporomandibular joint sounds. IEEE Trans on biomed Eng 47: 977–984.
- Djurdjanovic D, Ni J, Lee J (2002) Time-frequency based sensor fusion in the assessment and monitoring of machine performance degradation, Proceedings of 2002 ASME Int. Mechanical Eng. Congress and Exposition, paper number IMECE2002–32032.
- Eason K (1988) Information technology and organizational change, Taylor and Francis.
- Emmanouilidis C (2002) Evolutionary Multiobjective Feature Selection and ROC Analysis with Application to Industrial Machinery Fault Diagnosis, in Evolutionary Methods for Design Optimisation and Control. K. Giannakoglou, D. Tsahalis, J. Periaux, K. Papailiou and T. Fogarty (Eds.), CIMNE, ISBN:84-89925-97-6, Barcelona.
- Emmanouilidis C, MacIntyre J and Cox C (1998) Neurofuzzy Computing Aided Machine Fault Diagnosis, Proceedings of JCIS'98, The Fourth Joint Conference on Information Sciences, Research Triangle Park, North Carolina, USA, 24–28 October, 1998. Vol.1 pp. 207–210.
- Emmanouilidis C, Jantunen E, MacIntyre J (2006) Flexible Software for Condition Monitoring, Incorporating Novelty Detection and Diagnostics. Comput in Ind 57(6): 516–527.
- Fernandez O, Labib AW, Walmsley R, Petty DJ (2003) A decision support maintenance management system: development and implementation. Int J of Qual & Reliab Manag 20: 965–979.
- Fox P, *et al.* (2006) Will mobiles dream of electric sheep? Expectations of the new generation of mobile users: misfits with practice and research. in Proceedings of the international conference on mobile business (ICMB'06), Copenhagen, Denmark: IEEE.
- Garga A *et al.* (2001) Hybrid reasoning for prognostic learning in CBM systems, Aerospace Conference, IEEE Proceedings. 6: 2957–2969.
- Goodenow T, Hardman W, Karchnak M. (2000) Acoustic emissions in broadband vibration as an indicator of bearing stress. Proceedings of IEEE Aerospace Conference, 2000; 6: 95–122.L.D. Hall, J. Llinas (Eds.), Handbook of Sensor Fusion, CRC Press, 2000.
- Gunasekaran A, Khalil O, Rahman SM (2003) Knowledge and information technology management: Human and social perspective, Idea Group Publishing.
- Hall LD (1992) Mathematical techniques in Multi-Sensor Data Fusion, Artech House Inc.

- Hansen R, Hall D, Kurtz S (1994) New approach to the challenge of machinery prognostics, Proceedings of the International Gas Turbine and Aeroengine Congress and Exposition, American Society of Mechanical Engineers, June 13–16 1994; p.1–8.
- Hardman W, Hess A, Sheaffer (2000) A helicopter power-train diagnostics and prognostics demonstration, Proceedings of IEEE Aerospace Conference, 2000, 6: 355–366.
- Hurson AR, Pakzad S, Lin B (1994) Automated knowledge acquisition in a neural network based decision support system for incomplete database systems. *Microcomput in Civ Eng* 9(2): 129–143.
- IMS (2004) NSF I/UCRC Center for Intelligent Maintenance Systems, www.imscenter.net;
- Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition based maintenance. *Mech Syst and Signal Process* 20(7): 1483–1510.
- Jung B, Marquez AC (2006) Computers in Industry. *Comput in Ind* 57(6): 473–475
- Jeng MD (1997) Petri nets for modeling automated manufacturing systems with error recovery. *IEEE Trans on Robotics and Autom* 13: 752–760.
- Jones GJF (2005) Challenges and Opportunities of Context-Aware Information Access. in 2005 International Workshop on Ubiquitous Data Management (UDM 2005). Tokyo, Japan.
- Keen PGW (1981) Value analysis: justifying decision support systems. *Manag Inf Syst Q* 5(1): 1–15.
- Kemerait R (1987) New cepstral approach for prognostic maintenance of cyclic machinery. *IEEE SOUTHEASTCON*:256–262.
- Klein M, Methlie LB (1990) *Expert Systems: A Decision Support Approach*, Addison-Wesley, New York.
- Kleinbaum D (1994) *Logistic regression*. New York: Springer-Verlag.
- Koc MJ, Ni J, Lee J, Bandyopadhyay P, (2003) e-Manufacturing - trends and opportunities. *Int J of Adv Manuf Syst* 6(1): 29–46, 2003
- Krause A (2006) Context-aware mobile computing: learning context- dependent personal preferences from a wearable sensor array. *IEEE Trans on Mob Comput* 5(2): 113–127.
- Krotov V, Junglas I (2006) Mobile Technology as Enabler of Organizational Agility. in Proceedings of the International Conference on Mobile Business (ICMB'06). Copenhagen, Denmark.
- Kwon O (2006) The potential roles of context-aware computing technology in optimization-based intelligent decision-making. *Expert Syst with Appl* (31): 629–642.
- Labib AW (2006) Next generation maintenance systems: Towards the design of a self-maintenance machine, 2006 IEEE International Conference on Industrial Informatics, Integrating Manufacturing and Services Systems, 16–18 August, Singapore
- Laso-Ballesteros I (2006) Research perspectives on collaborative infrastructures for collaborative work environments. in 15th International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises. 2006. Manchester, U.K.
- Lee J (1995) Machine performance monitoring and proactive maintenance in computer-integrated manufacturing: review and perspective. *Int J of Comput Integr Manuf* 8: 370–380.
- Lee J (1996) Measurement of machine performance degradation using a neural network model. *Comput in Ind* 30: 193–209.
- Lee J and Ni J (2002) Infotronics agent for tether-free prognostics. Proceeding of AAAI Spring Symposium on Information Refinement and Revision for Decision Making: Modelling for Diagnostics, Prognostics, and Prediction, Stanford Univ., Palo Alto, CA, March 25–27, 2002.
- Lee J, Ni J, Djurdjanovic D, Qiu H and Liao H (2006) Intelligent prognostics tools and e-maintenance. *Comput in Ind* 57: 476–489.
- Levitt J (1997) *Maintenance Management, The Handbook of*. New York: Industrial Press.

- Liang E, Rodriguez R, and Hussein A (1988) Prognostics/diagnostics of mechanical equipment by neural network, *Neural Networks* 1(1): 33
- Liao H, Lin D, Qiu H, Banjevic D, Jardine A, Lee J (2005) A predictive tool for remaining useful life estimation of rotating machinery components, *ASME International 20th Biennial Conference on Mechanical Vibration and Noise*, Long Beach, CA, 24–28 September, 2005.
- Liberatore MT, Stylianou AC (1994) Using knowledge-based systems for strategic market assessment. *Inf and Manag* 27(4): 221–232.
- Liu J, Djurdjanovic D, Ni J, Lee J, (2004) Performance similarity based method for enhanced prediction of manufacturing process performance, *Proceedings of the 2004 ASME International Mechanical Engineering Congress and Exposition (IMECE)*, 2004, Paper No. IMECE2004-62246
- Liyanage JP (2006) A socio-technical perspective on Integrated operations for high-risk and complex industrial assets: Experience from North Sea oil and gas industry, *The 1st World Congress on Engineering Asset Management (WCEAM)*, Brisbane, Australia.
- Liyanage JP and Bjerkebak E (2007) Key Note paper: Use of advanced technologies and information solutions for North sea offshore assets: Ambitious changes and Socio-technical dimensions, *Journal of International Technology and Information Management (JITIM)*, pp 1–10.
- Liyanage JP and Langeland T (2007) Smart assets through digital capabilities, Mehdi Khosrow-Pour (ed.), Chapter accepted to the *Encyclopaedia of Information Science and Technology*, IGI publishers, USA.
- Liyanage JP, Herbert M, Harestad J (2006) Smart integrated e-operations for high-risk and technologically complex assets: Operational networks and collaborative partnerships in the digital environment, Wang, Y.C., et.al., (ed.), *Supply chain management: Issues in the new era of collaboration and competition*, Idea Group, USA, pp. 387–414.
- Lu YH, Ebert DS, Delp EJ (2006) Resource-driven content adaptation, in *Computational Imaging IV*. *Proceedings of the SPIE*, C.A. Bouman, E.L. Miller, and I. Pollak, Editors. SPIE. p. 167–177.
- Lugmayr A, Saarinen T, Tournut JP (2006) The Digital Aura -Ambient Mobile Computer Systems. in *14th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP'06)*. Montbeliard-Sochaux, France.
- Mann LJ (1983) *Maintenance Management*. Lexington, Massachusetts: Lexington Books.
- Manvi SS and Venkataram P (2004) Applications of agent technology in communications: a review. *Comput Commun* 27: 1493–1508.
- Marple S (1987) *Digital spectral analysis*. NJ: Prentice Hall.
- Marseguerra M, Zio E, Podofillini L (2002) Condition-Based Maintenance Optimization by Means of Genetic Algorithm and Monte Carlo Simulation. *Reliab Eng and Syst Saf* 77: 151–166.
- Mechefske CK and Wang Z (2001) Using fuzzy linguistics to select optimum maintenance and condition monitoring strategies. *Mech Syst and Signal Process* 15 (6): 1129–1140.
- Mijailovic V (2003) Probabilistic Method for Planning of Maintenance Activities of Substation Component. *Electr Power Syst Res* 64: 53–58.
- Muller A, Marquez AC, Iung B (2008) On the concept of e-maintenance: Review and current research. *Reliab Eng & Syst Saf* 93(8): 1165–1187.
- O'Grady MJ and Hare GMP (2005) Mobile Devices and Intelligent Agents - towards a new generation of applications and services. *Inf Sci* 171: 335–353.
- Panagiotakis S Alonistioti A (2006) Context-Aware Composition of Mobile Services. *IT Prof* 8(4): 38–43.
- Pandit S and Wu SM (1994) *Time series and system analysis with application*. FL: Krieger Publishing Co.

- Papadopoulos C (2005) Improving awareness in mobile CSCW. *IEEE Trans on Mob Comput* 5(10): 1331–1343.
- Parker BE, Nigor TM, Carley MP, Barron RL, Ward DG, Poor HV, Rock D, and DuBois TA (1993) Helicopter gearbox diagnostics and prognostics using vibration signature analysis, *Proceedings of the SPIE – The International Society for Optical Engineering*, 1965: 531–42
- Pham H, Suprasad A and Misra RB (1997) Availability and Mean Life Time Prediction of Multistage Degraded System with Partial Repairs. *Reliab Eng and Syst Saf* 56: 169–173.
- Radjou N (2002) The collaborative product life-cycle. Forrester Research, May 2002.
- Rao M, Theisen C, Luxhoj JT (1990) Intelligent system for airtraffic control. *Proceedings of the 5th IEEE International Symposium on Intelligent Control*, 2: 859–863.
- Ray A and Tangirala S (1996) Stochastic Modelling of Fatigue Crack Dynamic for On-Line Failure Prognostics. *IEEE Trans on Control Syst Technol* 4(4): 443–449.
- Reichard K, Van Dyke M, Maynard K (2000) Application of sensor fusion and signal classification techniques in a distributed machinery condition monitoring system. *Proceed of SPIE - The Int Soc for Opt Eng* 4051: 329–336.
- Roemer M, Kacprzyński G, Orsagh R (2001) Assessment of data and knowledge fusion strategies for prognostics and health management, *Proceedings of IEEE Aerospace Conference*, 6: 62979–62988.
- Saaty TL (1990) *The Analytic Hierarchy Process – Planning, priority setting, resource allocation*, RWS Publications, University of Pittsburgh.
- Sambamuthy V, Bharadwaj A, Grover V (2003) Shaping agility through digital options: reconceptualizing the role of information technology in contemporary firms. *MIS Quarterl* 27(2): 237–263.
- Santana M (1995) Managerial learning: A neglected dimension in decision support systems. *Proceeding: 28th Annual Hawaii International Conference on System Sciences* 4: 82–91.
- Schrunder CP, Gallertly JE, Bicheno JR (1994) A fuzzy, knowledge-based decision support tool for production operations management. *Expert Syst* 11 (1): 3–11.
- Seliger G, *et al.* (2002) Innovative processes and tools for disassembly, *Annals of CIRP* 51: 37–41.
- Shen, QP, Lo KK, Wang Q (1998) Priority setting in maintenance management: a modified multi-attribute approach using analytic hierarchy process. *Constr Manag and Econ* 16: 693–702.
- Sloan TW and Shanthikumar JG (2002) Using in-line equipment condition and yield information for maintenance scheduling and dispatching in semiconductor wafer fabs. *Scheduling and Logistics on Semiconductor Manufacturing. IIE Transactions (Inst of Ind Eng)*, 34(2): 191–209.
- Spaccapetra S, Al-Jadir L, Yu S (2005) Somebody, sometime, somewhere, something. in *Proceedings of the 2005 workshop on Ubiquitous Data Management*. 2005. Tokyo, Japan.
- Stoneham D (1998) *The Maintenance Management and Technology Handbook*. NY: Elsevier.
- Su L, Nolan M, DeMare G, Carey D (1999) Prognostics framework ‘for weapon systems health monitoring, *Proceedings of IEEE Systems Readiness Technology Conference, IEEE AUTOTESTCON '99*, 30 August–2 September 1999; p. 661–672.
- Swanson DC (2001) A General Prognostics tracking algorithm for predictive maintenance, *Proc. of the IEEE Aerospace Conference*, 2001, Vol.6, pp. 2971–2977.
- Tacer B, Loughlin P (1996) Time-frequency based classification. *SPIE Proceed* 42: 2697–2705.
- Tan CM and Wond CC (2006) Mobile Broadband Race: Friend or Foe? in *Proceedings of the International Conference on Mobile Business (ICMB'06)*. Copenhagen, Denmark: IEEE.

- Thurston M and Lebold M (2001) Open Standards for Condition Based Maintenance and Prognostic Systems, Pennsylvania State University, Applied Research Laboratory.
- Tse PW, Peng YH and Yam R (2001) Wavelet Analysis and Envelope Detection for Rolling Element Bearing Fault Diagnosis – Their Effectiveness and Flexibilities. *J of Vib and Acoust* 123(3): 303–310.
- Tu YL (1997) Integrated maintenance management system in a textile company. *Int J of Adv Manuf Technol* 13, pp. 241–258
- Umeda Y, Tomiyama T, Yoshikawa H (1995) A design methodology for self-maintenance machines, *ASME journal of mechanical design*, 117, September
- Vachtsevanos, G., Wang, P. (2001). Fault prognosis using dynamic wavelet neural networks, *Proceedings of the IEEE International Symposium on Intelligent Control 2001 (ISIC '01)*, pp 79–84.
- van Bunningen AH, Feng L, Apers PMG (2005) Context for Ubiquitous Data Management. in 2005 International Workshop on Ubiquitous Data Management (UDM 2005). 2005. Tokyo, Japan: IEEE Computer Society.
- Wang HZ (2002) A Survey of Maintenance Policies of Deteriorating Systems. *Eur J of Oper Res* 139: 469–489.
- Wang L, Chu J, Wu J, (2007) Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process. *Int J of Prod Econ* 107: 151–163.
- Want R and Pering T (2005) System challenges for ubiquitous & pervasive computing. in 27th International Conference on Software Engineering (ICSE 2005). St. Louis, Missouri, USA.
- Wickens CD (1992) *Engineering psychology and human performance*, Harper Collins.
- Wi-Fi Alliance (2007) <https://www.wi-fi.org>.
- Wikipedia (2007) <http://en.wikipedia.org>
- Wilson B, *et al.* (1999) Development of a modular in-situ oil analysis prognostic system, International Society of Logistics (SOLE) 1999 Symposium, Nevada, Las Vegas, 30 August–2 September, 1999.
- WiMax Forum (2007) <https://www.wimaxforum.org>.
- Yam RCM, Tse PW, Li L, Tu P (2001) Intelligent predictive decision support system for condition-based maintenance. *Int J of Adv Manuf Technol* 17: 383–391.
- Yamada A and Takata S (2002) Reliability improvement of industrial robots by optimizing operation plans based on deterioration evaluation, *Annals of the CIRP* 51: 319–322.
- Yang Z, Chang Q, Djurdjanovic D, Ni J, Lee J (2005) Maintenance Priority Assignment On-Line Production Information, Accepted for publication in the *Transaction of ASME, Journal of Manufacturing Science and Engineering*, Paper no. MANU-05-1073, 2005
- Yang Z, Chang Q, Djurdjanovic D, Ni J, Lee J (2006) Maintenance Scheduling in Manufacturing Systems Based on Predicted Machine Degradation”, Accepted for publication in *The Journal of Intelligent Manufacturing*.
- Yen G, Lin K (2000) Wavelet packet feature extraction for vibration monitoring. *IEEE Trans on Ind Electron* 47: 650–667.
- Yu R, Iung B, Panetto H (2003) A multi-agents based e-maintenance system with case-based reasoning decision support. *Eng Appl of Artif Intell* 16(4): 321–333.
- Zalubas E, *et al.* (1996) Shift and scale invariant detection. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Process* 5: 3637–3640.
- Zhu X (1996) Sensor-based condition monitoring and predictive maintenance: An integrated intelligent management support system. *Intelligent Systems in Accounting, Finance and Management*, 5, pp. 241–258
- Zimmermann A, Specht M, Lorenz A (2005) Personalization and Context Management. *User Model and User-Adapt Interact* 15(3-4): 275–302.

Maintainability and System Effectiveness

Maintainability and System Effectiveness

J. Knezevic

21.1 Introduction

According to Albert Einstein, “Everything that the human race has done and thought is concerned with the satisfaction of felt needs.” Einstein (1991). Thus, need for transporting, communicating, racing, entertaining, heating, navigating, and similar functions are daily manifested by the human race.

It is common practice to use the word system as a generic term for all solutions for satisfying human needs. System is a collection of mutually related components, selected and arranged after some distinct logical, scientific or instinctive method to perform at least one function with a measurable performance and attributes. Hence, the concept system became viable only when the measurable function is associated with a collection of components.

Although the felt needs cover a very large spectrum of solutions, the word engineering system is commonly used as a generic name for all of them. The most commonly used engineering systems in daily life are aeronautical and aerospace: agricultural, structural, chemical process and processing, civil engineering, electrical and electronic, metallurgical, nuclear, ocean, marine, nautical, petroleum, and similar.

The theoretical foundation of designing engineering systems are laws of nature, described through the laws of physics, like Newton’s laws of motion, Coulomb’s law of solid friction, Hooke’s laws of stress and strain, laws of thermodynamics, to name a few. As laws of nature are independent of time and the location in the universe, each individual system, of the same type, delivers identical function, performance and attributes, under identical conditions. As these expectations have not yet been proved wrong, the predictability of classical physics became known as determinism.

Any system starts its life by initial transition into the state in which it is able to deliver function, named “State of Functioning”, denoted as SoFu. If a system were

able to stay in that state for an indefinite period of time, the need for the further studies would not have existed. However, irrespective of how good the design is, how flawless the production process is, any engineering system during its life will reach a state in which it is unable to deliver function, performance or attributes caused by, some of the following factors:

- Inherent deficiencies of materials;
- Erroneous design decisions;
- Production and installation errors;
- Irreversible processes that take place in a system itself;
- Interaction of a system with its operational environment (natural and human);
- Planned execution of operational and maintenance tasks; and
- Insufficient operational and maintenance resources.

This new state of the system is commonly known as the “State of Failure”, denoted as SoFa.

In this new situation only two options are possible:

1. Do nothing and never satisfy felt need(s) again; or
2. Do something to restore the functionality of the system.

The process of “doing something” is known as the maintenance process. It can be viewed as the flow of maintenance tasks selected and performed by the user to retain or restore the functionality of the system during its operational process. Hence, successful completion of a maintenance task will cause the transition of the system to the SoFu. Experience teaches us that during the life of a system there will be many “failure” and “repair” events. Hence, from the point of view of “satisfying perceived needs” any engineering during its operational life will be fluctuating between SoFu and SoFa until its retirement is shown in Figure 21.1. The established pattern is termed the functionability profile because it maps states of the system during its operational life (Knezevic, 1996). Usually calendar time is used as the unit of operational time against which the profile is plotted.

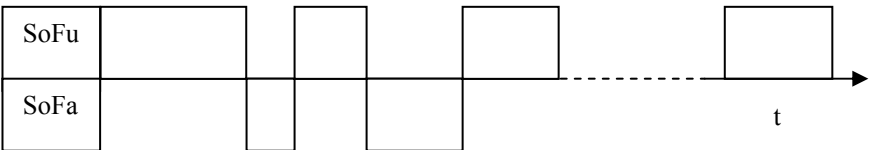


Figure 21.1. Functionability profile of an engineering system

It is extremely important for the user to have information about the function, performance, cost, safety, and other characteristics of the system under consideration at the beginning of its operational life. However, it is equally, or even more, important to have information about the characteristics which define the pattern of its functionability profile, as the main reason for the acquisition of any engineering system is the “satisfaction of felt needs”. Simply, an engineering

system is useful when, and only when, it performs the required function with expected performance. For example, a commercial aircraft makes money only when it is in the sky, flying the ticket-paid passengers between two destinations. The situation is the same with cars, cookers, computers, motorways, bridges, power stations, oil refineries, and so forth.

Thus, one of the main concerns of the users of any engineering system is the pattern of its functionability profile, with a specific emphasis on the proportion of the time during which the system under consideration will be available for the fulfilment of functionability. Clearly, the following two factors are chiefly responsible for its specific shape:

1. Inherent characteristics of a system, like reliability, maintainability, and supportability, which directly determine the frequency of the occurrence of failures, the complexity of maintenance tasks and the ease of the support of the tasks required, all of which are determined by the decisions made by the designers and constructors at the early stages of the system design.
2. Operational characteristics of a system, which are driven by the operational scenario, maintenance policy and the logistics support concept, determined by the each user of the each system, with the objective to manage the provision of the resources needed for the successful completion of all operation and maintenance tasks (Blanchard, 1969).

Consequently, any figure of merit that is used to define the effectiveness of any engineering system is uniquely defined by the shape of the functionability profile, resulting from the joint contributions made by the producer and users. Most frequently used figures of merit for the system effectiveness are availability, readiness, despatch reliability, mission reliability, and similar. In many situations, effectiveness figures of merit are combined with associated cost figures to form cost effectiveness figures of merit, like cost/seat/mile and similar.

Example 21.1: A quick look at the logbook of the first Boeing 747 owned by PanAm, registration number N747PA, clearly illustrates the interaction between the operation and maintenance processes during the 22 years of service. Thus, this particular aircraft has:

- Been airborne 80,000 h;
- Flown 37,000,000 miles;
- Carried 4,000,000 passengers;
- Made 40,000 take-offs and landings; and
- Consumed more than 271,000,000 gallons of fuel.

These are some statistics related to the SoFu, driven by the PanAm's business plan.. In order to meet the above given operational scenario, among many other resources consumed, it has:

- Gone through 2100 tyres;
- Used 350 brake systems;
- Been fitted with more than 125 engines;

- Had the passenger compartment and lavatories replaced four times;
- Had structural inspections for metal fatigue and corrosion which have needed more than 9800 individual X-ray frames of film; and
- Had the metal skin, on its superstructure, wings and belly replaced five times.

Replacement of the above-mentioned items and others, coupled with all other maintenance tasks performed during the 22 years of operation, accumulated 806,000 maintenance hours. Based on the above statistics, a large number of the system's operational and cost effectiveness figures of merit could be calculated. For example, it is easy to calculate that for every hour of "satisfying the needs" for flying the aircraft consumed 10 h of maintenance (Knezevic, 1996).

Example 21.2: Daily Mail, in the UK, reported on 13th December 1990, in the article entitled Warships 'wasting years stuck in dock' that Members of British Parliament's investigation revealed that "A frigate or destroyer spends 8 years of its average 22-year "life" under maintenance, and only half of the remaining 14 years would be spent at sea". This practically means that the frigate/destroyer of this type under this specific operational scenario is available to the Navy around 60% of the time when needed.

The rest of the chapter is organized as follows. Section 21.2 presents the concept of maintainability followed by maintainability analysis in Section 21.3. Maintainability measures are presented in Section 21.4 and the engineering methodology for predicting measures of system performance is provided in Section 21.5. The role of maintainability engineering management is outlined in Section 21.6 followed by concluding remarks in Section 21.7.

21.2 The Concept of Maintainability

Based on the above, it is not difficult to conclude that, from the point of view of the user, the shape of the functionability profile is one of the most important characteristics of the system, even more than the proportion of the time during which the system under consideration is available for the satisfaction of needs. For example, a commercial aircraft makes money only when it is flying, *i.e.*, transporting passengers between two destinations. The situation is the same with cars, cookers, computers, *etc.*

Designers, primarily of aerospace and military products, have been under huge pressure from users, in the last 30 years, to provide some information regarding the expected shape of the functionability profile, together with a recommended list of type and quantity of resources needed for its achievement.

As it is extremely important for the operators/users to know the functionability, durability and reliability characteristics of the system at the beginning of its operational life, it is equally, or even more, important for them to have the information regarding issues like:

- Which maintenance tasks should be performed?
- When the maintenance tasks should be performed?

- How difficult is it to perform a maintenance task?
- How safe is it to perform a maintenance task?
- How many people are required to perform a maintenance task?
- How much the restoration is going to cost?
- How long the system is going to be in a state of failure?
- What equipment is required? and
- What worker skills are needed to perform the prescribed activities?

In most cases the answers to these questions provided by designers/manufacturers are very basic and limited. For example, in the case of motor vehicles the answers cover no more than the list of maintenance activities which should be performed during regular service every 6,000 miles or so. All the above questions remain unanswered; the users are left to find the answers by themselves. The reason for this is the fact that, up to now, the main purpose and concern of designers is the achievement of function, whereas the ease of maintaining function by the users has been almost ignored. Traditionally it was the “problem” of the maintenance personnel, not the designers.

However, today the situation is gradually changing, thanks to aerospace and military customers who recognised the importance of the information of this type and who made it an equally desirable characteristic as a power, speed, weight and similar.

As none of the existing scientific or engineering disciplines were able to help designers and producers to provide an answer to the above question the need arose to form a new discipline. Maintainability Theory was created, defined as:

A scientific discipline which studies complexity, factors and resources related to the tasks needed to be performed by the user in order to maintain an engineering system in the State of Functioning, and works out methods for their quantification, assessment, prediction and improvement, (Knezevic, 1996).

Maintainability theory is rapidly growing in importance because of its considerable contribution towards the reduction of maintenance cost of engineering systems during its operational life. At the same time, in order to be used in daily practice, maintainability as a characteristic of engineering systems has to be defined. In technical literature several definitions for maintainability can be found. For example, MIL-STD-721B (1966) defines maintainability as: a characteristic of design and installation which is expressed as the probability that an item will be retained in or restored to a specified condition within a given period of elapsed time, when maintenance is performed in accordance with prescribed procedures and resources.

However, in this book the definition proposed by Knezevic (1996) is used:

Maintainability is the inherent characteristic of an engineering system related to its ability to be maintained in the State of Functioning by performing the required maintenance tasks as specified.

It is necessary to stress that maintainability theory could be a very powerful tool for engineers and managers to quantify and assess the ability of their systems to be maintained in SoFu during operational life.

21.2.1 Maintainability Impact on System Effectiveness

The majority of users state that they “need the equipment functionality as badly as they need safety, because they cannot tolerate having equipment out of operation”. There are several ways that designers can control that. One is to build items/systems that are extremely reliable, and consequently, costly. The second is to provide a system that, when it fails, is easy to restore. Thus, if everything is made highly reliable and everything is easy to repair, the producer has got a very efficient system that no one can afford to buy. Consequently, the question is how much a utility of the system is needed, and how much is one prepared to pay for it? For example, how important for the train operator is it to move train from the platform, when 1000 fare paying passengers expect to leave the gate at 6.25 am? Clearly the passengers are not interested what the problem is, or that it is designer's error, manufacturers, maintainers, operators or somebody else's problem. They are only interested in leaving at 6.25 am in order to arrive at their chosen destination at 7.30 am. Thus, if any problem develops, it needs to be rectified as soon as possible.

Consequently, maintainability is one of the main factors in achieving a high level of operational effectiveness, which in turn increases users or customers' satisfaction.

Example 21.3: The main objective of this example is to illustrate the impact of maintainability on operational effectiveness of motor vehicles. The inherent maintainability characteristics for several motor vehicles are given in Table 21.1. They clearly indicate the impact of the design decisions on the maintenance resources, frequency, and ultimately operational effectiveness.

Table 21.1. Maintenance intervals, durations and replacement time

Model	Major Service		Replacement time in hours						
	Interval (miles)	Duration (h)	Clutch	Exhaust	Rear damper	Head-lamp	Wind-screen	Front bumper	Alternator
Montego 1.6	12,000	2.6	3.9	1.2	1.5	0.4	2.5	1.0	0.6
Peugeot 205	12,000	1.8	3.7	1.0	1.4	0.4	2.0	0.3	0.5
Astra GTE	9,000	1.4	1.2	0.9	0.6	0.6	0.2	0.8	0.6
Jetta 1.8	10,000	2.0	2.9	0.9	0.5	0.4	0.7	0.4	0.5
Toyota C.	10,000	2.0	3.9	1.3	0.8	0.4	2.9	0.8	0.7
Lada 1500	6,000	3.6	3.2	1.8	0.9	0.2	0.5	0.5	0.7
Cavalier	9,000	1.3	1.2	0.9	0.6	0.7	1.3	0.6	0.5
Golf 1.6	10,000	2.0	3.3	0.9	0.6	0.7	1.3	0.6	0.5
Sierra 1.6	12,000	2.4	2.0	0.6	0.4	0.4	2.1	0.4	0.4
Nisan Micra	6,000	2.8	3.3	0.7	1.6	0.2	1.8	0.6	0.6
Renault 5	12,000	3.6	4.4	1.3	0.4	0.4	0.4	0.4	0.4
Alpha 33	12,000	3.0	4.4	0.5	0.4	0.2	1.8	0.4	0.3

Based on the above data for the specific operational scenario, where it was assumed that during a 3-year period the total mileage covered by each type of

motor vehicle was 75,000, the total hours spent on maintaining their functionability, by the users, is given in Table 21.2 together with the operational effectiveness achieved.

Table 21.2. Impact of maintainability on functionability

Car MTIM	No of Services	MTIM ^p (h)	MTIM ^c	MTIM	Functionability
Montego 1.6	6	15.6	17.2	38.2	0.9927
Peugeot 205	6	10.8	14.1	24.9	0.9945
Astra GTE	8	11.2	14.1	25.3	0.9955
Jetta 1.8	7	14.0	11.1	25.1	0.9944
Toyota C.	7	14.0	17.0	31.0	0.9931
Lada 1500	12	43.2	18.3	61.5	0.9863
Cavalier	8	9.1	8.9	18.0	0.9960
Golf 1.6	7	14.0	15.9	29.9	0.9934
Sierra 1.6	6	14.4	9.9	24.3	0.9946
Nisan Micra	12	33.6	13.8	45.4	0.9895
Renault 5	6	21.6	17.0	38.6	0.9914
Alpha 33	6	18.0	15.7	33.7	0.0025

In Table 21.2, MTIM^p represents the mean cumulative time in maintenance caused by the execution of preventive maintenance tasks (services), MTIM^c stands for the corresponding time caused by the demand for the execution of corrective maintenance tasks, and MTIM represents the mean total time in maintenance obtained as a sum of the two ($MTIM = MTIM^p + MTIM^c$).

Example 21.4: Maintenance troubleshooting is another area to be considered under the maintainability heading. For the airlines this is usually only about 1 h at the gate prior to its departure to the next destination, whereas for a racing car or weapon system this is usually a few minutes. An easily manageable device is needed for the diagnostic of all different modules in order to determine their state and identify the failed one within it. Practice shows that false removals cost about the same as an actual failure when the component under investigation is removed and replaced. Reducing this would be a big cost saver. Devices with such capabilities have been developed in the aerospace industry as a result of maintainability studies and research. For example, the design of the Boeing 777 includes “On-Board Maintenance System” with the objective to assist the airlines with a more cost-effective and time-responsive device to avoid expensive gate delays and flight cancellations [Proctor, Journal “Aviation Week & Space Technology”]. For similar purposes the Flight Control Division of Wright Laboratory, USAR, has developed, fault detection/isolation system for F-16 aircraft, which allows maintainers – novice as well as expert – to find a failed component.

With the older fleets, both in the military and commercial sectors there is a great need for easier detection of corrosion. When these aircraft were built, they were designed for a certain life cycle, not for the extended service imposed on them. As the number of flying hours of an aircraft increase, the chance of corrosion and structural fatigue increases. One of the objectives of maintainability is the development of the system for detection and identification of failures before they make aircraft safety critical.

One of the common perceptions is that maintainability is simply the ability to reach a component to change it. However, that is only a small aspect. Maintainability is actually just one dimension of system design and a system's maintenance management policy. For example, it could be required from the designer that only three screws are acceptable on a certain partition panel in order to get speedy access inside. However, this request has to be placed into a larger context and it becomes a trade-off. If the item behind that panel needs to be checked once in every 5–6 years, or say 50,000 miles, it does not make much sense to concentrate much intellectual effort and spend project money on quick access. Thus, a lot of fasteners and connectors could be tolerated and the item may not be quickly accessible, but all of that has to be traded off against the cost and operational effectiveness of the system.

Additionally, decision makers have to be aware of the environment in which maintainers operate. It is much easier to maintain an item on the bench than at the airport gate, war theatre, busy morning traffic, or any other result-oriented and schedule-driven environment. Thus, the trade off process has to take into account the operational environment and the significance of the consequences if the task is not completed satisfactorily, when the trade off is made. According to Hessburg, the chief mechanic of new airplanes from Boeing:

“Maintenance managers want a clean gate, their report card in line maintenance based on having a clean gate and not having pigeons roosting on the airplane's fin. So it is necessary to try to influence the design that way, and say “here's what mechanics have to do at the gate” [May 1994, Aviation Equipment Maintenance].

The majority of users are currently showing concern over the competitive advantage that maintainability and maintenance can provide to a company. To illustrate the economic importance of maintenance, a recent study of engineering maintenance practices show that:

- United States airlines spend 9 billion dollars, approximately 11% of their operating cost, on maintenance [Journal, Aviation Week & Space Technology].
- Military sector has even higher concern for the maintenance cost that accounts for about 30% of the life cycle cost of weapon system. In 1987/88 the Royal Air Force spent around 1.9 billion pounds on the maintenance of aircraft and equipment.
- British manufacturing industry, according to the report produced by the Department of Trade and Industry, spends 3.7% of annual sales value each year on maintaining direct production system. Translating the above percentage to the sum of money spent in UK industry on maintenance it amounts to 8.0 billion pounds in a year (1988).

21.2.2 Maintainability Impact on Safety

Finally, performance of any maintenance tasks is related to an associated risk, both in terms of the non-correctly performed specific maintenance task, and the consequences of performing the task on the other item of the system, *i.e.*, possibly of inducing a failure on the system while doing maintenance.

Example 21.5: The Airbus A320, owned by Excalibur Airways, performed an un-commanded roll to the right due to loss of spoiler control just after take-off from Gatwick Airport in London, UK, in August 1993. A report released in February 1994 by the Air Accidents Investigation Branch (AAIB) stated “the emergency rose, not from any mechanical malfunction, but from a complex chain of human errors by the maintenance crew and by both pilots.” Apparently, during the flap change, maintenance did not comply with the maintenance manual. The spoilers were placed in maintenance mode and the collars and flags were not fitted. Also, the reinstatement and functional check of the spoilers after flap installation were not carried out. In addition, the pilots failed to notice during the independent functional check of the flight controls that spoilers two through five on the right wing did not respond to the right roll commands. The AAIB made 14 safety recommendations to the Civil Aviation Authority including formally reminding technicians of their responsibility to ensure all work is carried out in compliance with the maintenance manual and no work otherwise is to be certified. It also recommended that Airbus amend the A320 maintenance manuals concerning flap removal, and that the flap refitting and spoiler de-activation chapters include specific warnings to reinstate and function the spoilers after deactivation [April 1995, Aviation Equipment Maintenance].

Example 21.6: In the article entitled Hangar error published in January 1992, in the journal Aerospace, the following three, maintenance related, accidents have been exposed:

1. At Chicago airport the DC-10 rolled onto its back after take-off and crashed into a caravan park, killing all 272 on board and 2 people on the ground. The cause of the accident was an engine separation due to a fatigue in a cracked engine mounting which resulted from an improper fork lifting short cut. After the accident, other DC-10 maintenance engineers said they had used the same short cut. One had actually heard a sharp cracking noise from the structure but had not dared to report it.
3. The total engine failure of a TriStar was caused by the incorrect insertion of oil chip-detectors, with O-ring seals missing. There had been 12 previous similar occurrences in the same airline, seven leading to unscheduled landings. This was a classic case of boredom and complacency in the hangar.
4. The total engine failure of a 767 was caused by misreading a dipstick in gallons instead of litres. Many other cases could be added to the list of maintenance related accidents which had very nearly happened before but were not reported. It is very difficult to admit to the boss that the litres had been misread for gallons.

Analysis of major civil aviation accidents resulting from errors made during the execution of maintenance tasks shows that between 1981 and 1985 there were 19 maintenance related failures which in total claimed 923 lives, detailed in Table 21.3. The biggest accident took place on 12 August 1985, when a JAL owned Boeing 747 decompressed from fatigue because of an improperly repaired bulkhead, killing 520 people.

Table 21.3. Maintenance related failures 1981–1985

Date	Carrier	Aircraft	Fatalities	Circumstances
26 June 81	Dan-Air	747	3	Cargo door separated and wrapped around tail plane
2 Aug 81	Far-Eastern	737	110	Structure failure of pressure cabin belly. Corrosion and fatigue
22 Sept 81	Eastern	TriStar	-	Uncontained failure No 2 engine. Extensive damage
13 Sept 82	Spantax	DC-10	51	Over-run and fire after nose-wheel tyre failure
16 April 83	AirLiberia	747	17	Engine failure
29 April 83	CAN	Caravelle	8	Engine failure
5 May 83	Eastern	TriStar	-	Total power loss - "O" rings missing
3 June 83	AirCanada	DC-9	28	Emergency landing after in-light fire in lavatory electrical system
11 Oct 83	AirIllonis	747	9	Total loss after despatch with unserviceable generator
14 Dec 83	Tampa	707	6	Uncontained engine failure
22 March 84	Pacific Western	737	-	Uncontained engine failure causing fire, write-off and injuries
13 June 84	Austrian	DC-9	-	Uncontained engine failure, fire, hydraulic failure
30 Aug 84	Cameroon	737	2	Uncontained engine failure causing fuel tank fire
18 Sept 84	AECA	DC-8-55	4	Power loss on take-off
21 Jan 85	TPI	Electra	71	Leading edge hatch opened
12 Aug 85	JAL	747	520	Decompressed of fatigue and improperly repaired bulkhead
22 Aug 85	BA	737	55	Uncontained engine failure and fire on take-off
15 Aug 85	Alyemda	707	3	Elevator control lost, emergency landing
6 Sept 85	Midwest	DC-9	36	Control lost after uncontained engine failure

The same analysis shows that during 1986–1990 there were 27 maintenance related failures claiming 190 lives. The most tragic of them was crash of a United owned DC-10 in 1989, when the fatigue of fan disc of the second engine caused complete hydraulic and flight control failure, and loss of 111 lives. Details are shown in Table 21.4.

Table 21.4. Maintenance related failures 1986–1990

Date	Carrier	Aircraft	Fatalities	Circumstances
10 March 86	BA	747	-	Rudder attachment failed
25 March 87	AA	DC-10	-	Cabin smoke, five injured
7 June 87	Transweed	Caravelle	-	Elevator fault on take off
24 July 87	Bouraq	747	-	Elevator locked on take-off (6)
5 Dec 87	USAir	737-200	-	No 2 engine separated (fatigue)
2 Feb 88	Express	Saab 340	-	Uncontained engine failure and fire
13 Feb 88	SAS	DC-10		Pitched, autopilot fault, ten injured
14 April 88	Piedmont	F28		Uncontained engine failure, decompression
29 April 88	Aloha	737	1	Cabin roof separation. Corrosion and fatigue
6 July 88	LAS Colombia	CL-44	3	Uncontained engine failure causing loss of control
11 Sept 88	BA	747-100		Flap fracture needing full aileron
8 Jan 89	BMA	737	55	Fan blade failure at top of climb, fatigue
20 Jan 89	Piedmont	737	-	No 2 engine separated after take-off
24 Feb 89	United	747	9	Cargo door separated, worn latch suspected
9 March 89	USAir	737	1	Cabin decompression
18 March 89	Evergreen	DC-9	2	Cargo door separation
2 June 89	Qantas	747	-	Autopilot excursion 31,000ft
9 June 89	Dan-Air	737	-	Fan blade failure at top of climb, fatigue
11 June 89	BMA	737	-	Fan blade failure at top of climb, fatigue
19 July 89	United	DC-10	111	No 2 fan disc fatigue causing complete hydraulic and flight control failure
16 Aug 89	Airfast	747	-	Cargo door 7000ft
4 Jan 90	Northwest	727	-	Engine fell off (blue ice)
7 May 90	Air-India	747	-	Engine fell off thrust reverse
11 May 90	Philippines	737	8	Fuel tank explosion. Faulty float switch
10 June 90	BA	One-Eleven	-	Windscreen blew out, injuring Captain. Wrong retainer bolts
15 June 90	TWA	TriStar	-	Cabin smoke on take-off, 72 incapacitated
11 Dec 90	AirCanada	TriStar	-	Aft pressure bulkhead decompression

21.2.3 Undesirable Maintainability Practices

Several real life examples are cited here in order to illustrate some of the undesirable maintainability decisions made in the past, which have caused considerable problems to the users.

Example 21.7: The engine starting system on Hunter aircraft: as the rapid starting of the heavy and large, Avon 200 engine was a dominant operational characteristic, the designers concentrated on a small turbine starter powered by iso-propyl-nitrate. Its high inertia forced the turbine to work at the peak of its performance. In the case of overspeed it could have damaged the engine, which would certainly have been catastrophic in the air. Consequently, the design was reviewed and a relay system introduced in order to shut down the start cycle if the starter turbine had not disengaged by 1600 rpm. This was good design decision, especially from the safety point of view, but very little consideration had been made regarding the reliability and maintainability issues. Hence, due to very high failure rate of the redesigned system, the aircraft functionability was drastically reduced, especially due to fact that it could not be changed in site, unless the mechanic of the day had “3 m long arms”. Consequently, the engine had to come out. Unfortunately, to achieve this, the back of the aircraft had to be removed. To achieve this activity the engine and flying control connectors had to be disconnected. Final results in the field were: 40 man-hours to change a relay, of which approximately 5 min was spent actually changing the relay itself. On top of that, every time the squadron went on detachment, the maintainers had to take along a full set of bulky support equipment to satisfy the inevitable need to change a few relays. [Source: Air Commodore O. Truelove, RAF, presentation at Exeter University, UK, 1989].

Example 21.8: The engine change on Harrier GR3. In order to perform this task the wing of aircraft must be removed. In order to achieve this it is necessary to disconnect a variety of control systems. The total task requires 24 h of elapsed time involving an assortment of heavy and bulky support equipment. [Source: Air Commodore O. Truelove, RAF, presentation at Exeter University, UK, 1989].

Example 21.9: The Times, on 11th February 1995 reported the following story: Comfortable Renault 25 TX, with nearly 75,000 on the clock, had been almost completely trouble-free during its life. However, the alarm bells rang only mildly when the heater stopped working and the temperature gauge refused to move, but then after 10 min driving sprang straight into the red. The technician at the Renault garage sounded mournfully like a doctor diagnosing a long, painful and exotic illness. “The heating matrix has gone, about the worst thing that could have happened. This is most unusual. Jolly bad luck.” The heating matrix is an oblong metal construction 30 cm by 15 cm by 5 cm, shaped like a small radiator, the main function of which is to provide warm air to heat the car. They are supposed never to go wrong, so manufacturers snuggle them deep in the car where they can remain untouched until the vehicle is scrapped. However, when they do fail, trouble and cost follow. The price of the heating matrix itself was £57.50 (FY1995). The total cost of the replacement, however, was £553.30, including VAT. This is because it

took 10.5 h work to get the old one out and put the new one in. The mechanics had to dismantle virtually the whole dashboard, remove most of its innards and, key-hole-surgery-style, negotiate the matrix out through the glove compartment. The work took a couple of days and the user had no use of the vehicle while the major surgery progressed. Renault's head office in Britain confirmed that 10.5 h was the correct amount of labour time needed to replace the matrix on that particular model. However, Renault pointed out that on their latest model, the Laguna, as a result of design change, the same item could be replaced within 1.5 h. The matrix is now accessible through the engine rather than the glove compartment.

21.2.4 Desirable Maintainability Practices

Certainly there are many more desirable maintainability practices, where efforts have been made at the design stage with the objective of making positive contributions towards the ease, accuracy and safety of maintaining the functionality of the system by the user during the utilization phase.

Example 21.10: During the course of most Formula 1 races, cars make at least one mid-race stop at their pits in order to change tires. The outcome of this maintenance task can occasionally mean the difference between first and second place. Consequently, in order to reduce the time spend in the pit to the minimum, the wheels of F1 cars are designed in a such way that only one central wheel nut provides a sufficient force for their attachments to the hub. Typical replacement times for all four wheels are given in Table 21.5.

Table 21.5. Tire change times, 1993 British Grand Prix

Team	Driver	Time (s)
McLaren	A. Senna	5.11
Benetton	M. Schumacher	5.50
Ligier	M. Brundle	6.75
Williams	D. Hill	7.61
Williams	A. Prost	8.02
Lotus	A. Zanardi	9.21

The above task requires 15 mechanics, 3 to remove and replace each wheel, 2 on quick-lift jacks, and the chief mechanic who holds a board in front of the car with signs "Brakes on/Go". These may be joined by another mechanic to steady the car.

The situation is similar with all other items, as illustrated in Table 21.6.

Table 21.6. Average replacement times in minutes

Item	Replacement time (min)
Engine	60
Gearbox	30
Four shock absorbers	12
Pedal box, seat and harness	10

Example 21.11: A complete operational turnaround for a SAAB-Gripen fighter aircraft, in the Swedish Air Force¹, including refuelling, reloading the gun, mounting six air-to-air missiles and making an inspection, can be performed with minimum equipment in less than 10 min by five conscripts under supervision of one technician. No tools are required to open and close the service panels, which are at a comfortable working height. All lights, indicators and switches needed during the turnaround are in the same area of the aircraft, together with the connections for fuel and communication with the pilot.

Example 21.12: GE Aerospace² transportable solid-state radar FPS-17 provides a system functionality of 0.996. This is achieved through huge reliable solid-state components, continuous automatic performance monitoring and fault isolation, and a mean time to repair of less than 30 min.

The situation is very similar to a Tactical Solid-State Radar AN/TPS-59. The reduction of maintenance costs (depot repair) is achieved through design improvements, like:

1. All printed wire boards are plug-in;
2. All integrated circuits are plug-in;
3. Ninety percent of all electrical connections are implemented by plug-in connectors or screws and lugs;
4. All printed wire boards employ solder masking to prevent solder shorts, and silk screening to prevent component misplacement; and
5. Continuous on-line automatic performance monitoring, off-line fault location, permit maintenance by medium-skill-level personnel during operations.

¹ Source: The Gripen Logistics Concept, publication 950601, Saab Military Aircraft, 1995

² Source: Manager-Marketing, GE Aerospace, Radar Systems Department, Syracuse, N.Y. (USA).

21.3 Maintainability Analysis

How long is the maintenance task going to last?

In order to answer the above question and explain the physical meaning of maintainability, let us establish the link between the complexity of the maintenance task and the duration of the maintenance task, DMT, expressed in some of the time-based units (seconds, minutes, *etc.*). Hence, maintainability can be graphically presented as shown in Figure 21.2, where DMT represents the elapsed time needed for successful completion of a specified maintenance task (Knezevic, 1996).

The complexity of the maintenance task, even though an unknown value, is identical for all executions of the identical maintenance tasks on the identical systems under consideration; therefore, there is no need for assigning a numerical value to it.

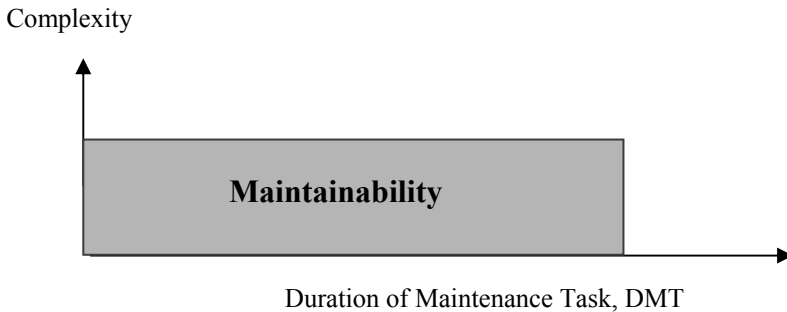


Figure 21.2. The DMT approach to maintainability

In spite of the fact that Figure 21.2 represents only an illustrative attempt at defining the meaning of maintainability, it also suggests that the ability of restoring functionality by performing a specified maintenance task could be numerically expressed by the indicated area. This means that maintainability is indirectly proportional to the area considered, *i.e.*, a less complex maintenance task performed on the system will cover a smaller area, and *vice versa*. It is necessary to stress that the size of the area considered mainly depends on the decisions taken during the design phase. In a sense, the order of magnitude of the length of the elapsed time required for the completion of a specific maintenance task (5 min, 5 h, or 2 days) could only be taken at very early stage of design process. Consequently, designers and constructors decisions taken during the early stages of the design related to the complexity of the maintenance task, manifested through:

- Accessibility of the items;
- Safety of the restoration;
- Troubleshooting procedure;
- Amount of testability built in;
- Physical location of the item; and

- Requirements for the maintenance support resources (material, facilities, spares, tools, equipment, trained personnel and similar).

These are prime drivers of maintainability characteristics. No amount of maintenance management procedures, quality controls and leadership skills can improve maintainability characteristics in the operational phase of an engineering system.

Thus, the maintainability could be quantitatively expressed through the length of elapsed time DMT during which the specified maintenance task is performed to the item under consideration and specified support resources used. The question that immediately arises here is: *what is the nature of DMT?* In other words, is DMT constant for each execution of the maintenance task considered or does it differ from trial to trial?

As the system under consideration physically exists only through copies made, the maintenance task exists only through its physical executions. Thus, the answer will depend on lengths of the elapsed time of each maintenance trial. In spite of the fact that each maintenance task consists of the specified activities that are performed in the specified sequence, the elapsed time needed for the execution of all of them might differ from trial to trial.

In order to illustrate this point, in Table 21.7 elapsed time needed for the replacement of the wheel on a small family car by a group of second year students from the School of Engineering of Exeter University in UK are given. Thus, the objective of this task is to restore functionality of a faulty tyre by replacing wheel and tyre assembly with functional one. The list of specified activities that have to be performed in a sequence is shown in Table 21.8.

Table 21.7. Length of time taken by a student to replace a car wheel, in seconds

Student	1	2	3	4	5	6	7	8	9	10
Time	230	259	442	286	397	365	332	279	321	351

Table 21.8. List of consisting maintenance activities

Order number	Activity
1	Remove spare wheel from a car boot
2	Take off wheel trim
3	Loosen all four bolts on existing wheel
4	Position and secure jack
5	Raise car
6	Remove bolts and take off the wheel
7	Replace the wheel and tighten bolts by hand
8	Lower jack
9	Tighten all four bolts
10	Install the wheel trim
11	Place the old wheel and jack in boot.

Maintenance tasks, like this one, are specified in the user manual that is delivered to the user together with the car, at the beginning of the operation of the system. Also, all maintenance resources needed for the successful completion of the task considered to be performed by the user have been provided by the manufacturer of the car to the user as a part of overall package. The list of resources needed for the task analysed is given in Table 21.9.

Thus, ten students performed this task individually on the same car following a list of specified activities that were to be performed in sequence. The tools required for execution of this task were laid out beside the wheel to be changed.

Table 21.9. List of required maintenance resources

Resources category	Specific resource
Personnel	Driver, no training required
Supply support	Spare wheel with a tyre
Equipment	Mechanical jack
Tool	Spanner 19mm
Facilities	Existing
Data	Tyre pressure
Manual	User's manual
Computer	Not required

In setting the task, an attempt was made to minimise the effect of various external factors. The task was performed in a garage in order to achieve stable environmental conditions. All participants were engineering students; an attempt was made to select a group with a similar mental approach, minimising personal factors. However, elapsed time differences indicate the individual variability of the skill levels, motivation, experience physical strength and similar issues. Generallyspeacking, if the elapsed time of several trials of a specific maintenance task is analysed it could be seen that the first task will be completed at the instant denoted by dmt_1 , second attempt at instant dmt_2 and say the n th will be executed by instant dmt_n , (Figure 21.3).

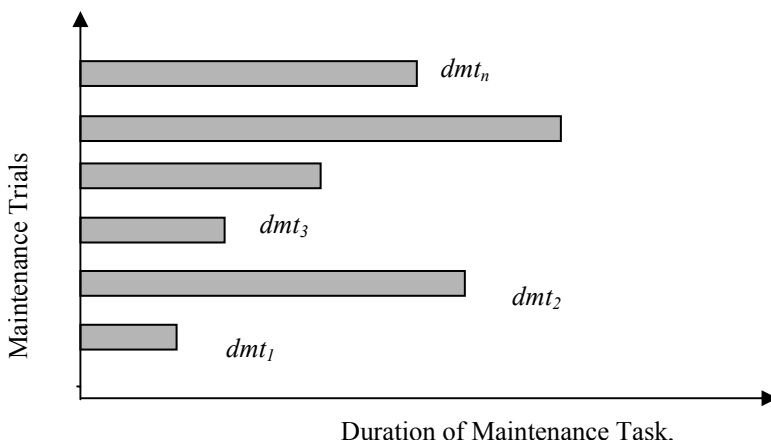


Figure 21.3. Maintainability pattern of a specific maintenance task formed by several trials

The above illustration only confirms what everyone familiar with the maintenance of engineering systems already knows: the execution of each trial of a specific maintenance task will be completed after a different interval of elapsed time. Thus, the length of elapsed time needed for the completion of each maintenance task is a specific characteristic of each trial.

The question that naturally arises here is: why are different lengths of an elapsed time needed for the execution of the identical maintenance tasks? In order to provide the answer to this question it is necessary to analyse factors that are responsible for that. According to Knezevic (1996), the following three groups are the most influential:

- Personal factors that represent the influence of the skill, motivation, experience, attitude, physical ability, eyesight, self-discipline, training, responsibility, patriotism and other similar characteristics related to the personnel involved;
- Conditional factors that represent the influence of the operating environment and the consequences of the failure to the physical condition, geometry, and shape of the item under restoration; and
- Environmental which represent the influence of factors such as temperature, humidity, noise, lighting, vibration, time of the day, time of the year, wind, noise, and similar to the maintenance personnel during the restoration.

Thus, the different lengths of elapsed time for the execution of each individual trial of the maintenance task considered are the result of the influence of the combination of the above-mentioned factors.

Consequently, the nature of the parameter DMT for the maintenance task also depends on the variability of those parameters. Therefore, the relationship between the influential factors and parameter DMT could be expressed by the following equation:

$$\text{DMT} = f(\text{personal, conditional and environmental factors})$$

Analysing the above expression it could be said that as a result of the large number of influential parameters in each group on one hand and their variability on the other, it is impossible to find the rule that would deterministically describe this very complex relation denoted with "f". The only way forward in the attempt to quantify maintainability is to call upon probability theory that offers a "mechanism" for the description of very complex relationships identified by the above expression.

In conclusion it could be said that it is impossible to give a deterministic answer regarding the instant of operating time when the transition from the SoFa to the SoFu will occur for any individual execution of the maintenance task under consideration. It is only possible to assign a certain probability that it will happen at a certain instant of maintenance time or that a certain percentage of trials will or will not be completed by the specific instant of elapsed time.

21.3.1 Measures of Maintainability

The most frequently used characteristics of DMT are (Knezevic, 1993) :

- Maintainability Function;
- Percentual Duration of Maintenance Task; and
- Expected/Mean Duration of Maintenance Task.

Each the measures are briefly described below.

21.3.1.1 Maintainability Function

This function, denoted as $M(t)$, represents the probability that the maintenance task considered will be successfully completed before or at the specified moment of maintenance, elapsed time t , thus:

$$M(t) = P(DMT \leq t) = \int_0^t m(\tau) d\tau \quad (21.1)$$

where $m(t)$ is the maintainability density function of DMT . According to Knezevic (1993), the most frequently used theoretical probability distributions to describe maintenance tasks related to the engineering systems are given Table 21.10.

Table 21.10. Mathematical expressions for maintainability function

Distribution	Expression	Range
Exponential	$1 - \exp(-t/A_m)$	$t \geq 0$
Normal	$\Phi[(t - A_m)/B_m]$	$-\infty \leq t \leq +\infty$
Lognormal	$\Phi[(\ln(t - C_m) - A_m)/B_m]$	$t \geq C_m, C_m \geq 0$
Weibull	$1 - \exp[-(t - C_m)/(A_m - C_m)]^{B_m}$	$t \geq C_m, C_m \geq 0$

In Table 21.10, where: A_m , B_m , and C_m are scale, shape and source parameters of probability distribution, and Φ is the standard Laplace function, value of which could be easily found in literature.

It should be noted that in the case of normal distribution the probability function exists from $-\infty$ so it may have a significant value at $t = 0$. Since negative time is meaningless in maintainability analysis, great care should be used in manipulating this model. Generally speaking, it could be said the normal probability distribution could be used to model the duration of the maintenance task in all cases where $A_m > 3 \times B_m$, as only in these cases can the numerical value of $M(t)$, at $t = 0$, can be considered negligible. In all other cases it could be possible to calculate the probability of completing the task before it began.

21.3.1.2 Percentual Duration of Maintenance Task DMT_p

This maintainability measure, denoted as DMT_p , represents the duration of maintenance task by which a given percentage of maintenance tasks considered will be successfully completed. It is the abscissa of the point whose ordinate presents a given percentage of a task completion. Analytically, DMT_p can be represented as

$$DMT_p = t, \text{ for which, } M(t) = P(DMT \leq t) = \int_0^t m(t)dt = p$$

The most frequently used value for the DMT_p is DMT_{90} , which presents the duration of the restoration time by which 90% of all executions of the maintenance task considered will be completed, thus:

$$DMT_{90} = t, \text{ for which, } M(t) = P(DMT \leq t) = \int_0^t m(t)dt = 0.9$$

It is worth noting that in military oriented literature and defence contracts, the numerical value of DMT_p is referred as “Maximum Repair Time” and it is denoted as M_{max} . Most frequently used value for M_{max} is 95-percentile value, thus $M_{max} = DMT_{95}$. This value is very beneficial for the planning purposes in “peace” time.

However, another potentially useful value of DMT_p could be “Minimum Repair Time” denoted as M_{min} . Suggested value for M_{min} is 10-percentile value, thus $M_{min} = DMT_{10}$. This value could be very beneficial for “war” or any other “competition” situations where it is crucial to make the decision whether or not to attempt to execute the task before the undesirable event is expected to happen.

21.3.1.3 Expected Duration of Maintenance Task

This maintainability measure, denoted as EDMT (also known as the Mean Duration of Maintenance Task), represents the expectation of the random variable. DMT can be used for calculation of this characteristic of maintenance process, thus:

$$E(DMT) = MDMT = \int_0^{\infty} t \times m(t)dt$$

The above characteristic could also be expressed in the following way:

$$E(DMT) = \int_0^{\infty} [1 - M(t)] dt$$

which represents the area below the function which is complementary to the maintainability function.

Analytical expressions for the Mean Duration of Maintenance Task, $MDMT = E(DMT)$, for well-known distributions, are given in Table 21.11.

Table 21.11. Analytical expressions for MDMT

Distribution	Expression
Exponential	A_m
Normal	A_m
Lognormal	$\exp(A_m + \frac{1}{2} B_m^2)$
Weibull	$A_m \times \Gamma(1 + 1/B_m)$

In Table 21.11 Γ is symbol for the well known Gamma function, whose numerical values could be found in reliability/maintainability literature.

Example 21.13: for the maintenance task, whose restoration time could be modelled by the Weibull distribution with parameters $A_m=29$, $B_m=2.9$ and $C_m=0$., determine: 1. the probability that the system will be restored in 20 min, 2. the time up to which 20% and 95% of the tasks will be successfully completed, 3. mean duration of maintenance task, MDMT.

1. Making use of Table 21.10, the maintainability function for this particular task is modelled by the following expression:

$$M(20) = 1 - \exp - \left[\frac{(20 - 0)}{(29 - 0)} \right]^{2.9} = 0.288$$

What is the probability that it will be restored in 35 min?

$$M(35) = 1 - \exp - \left[\frac{(35 - 0)}{(29 - 0)} \right]^{2.9} = 0.82$$

2. The TTR_p time represents the restoration time by which a given percentage of a maintenance task will be completed. For the Weibull distribution this can be calculated using the following equation:

$$t = A_m [-\ln(1 - M(t))]^{1/B_m}$$

$$TTR_{20} = 29 [-\ln(1 - 0.2)]^{1/2.9} = 17.29 \text{ min}$$

$$TTR_{95} = 29[-\ln(1 - 0.95)^{1/2.9}] = 42.33 \text{ min}$$

3. Expected restoration time:

This measure is the expected value of the random variable DMT and is also termed the Mean Duration of maintenance task (MDMT). It is calculated

using expression $E(TTR) = MTTR = \int_0^{\infty} [1 - M(t)] dt$. For the Weibull probability distribution, the numerical value for $E(TTR) = MTTR$, will be:

$$E(TTR) = MTTR = 29 \times \Gamma\left(1 - \frac{1}{2.9}\right) = 29 \times 0.892 = 25.87 \text{ min}$$

The numerical value for $\Gamma\left(1 - \frac{1}{2.9}\right) = 0.892$ was obtained from Table T_2 (Knezevic, 1996).

21.3.2 Maintenance Labour-Hour Factors

The maintainability measures, covered thus far relate to the elapsed maintenance times only. Although elapsed times are extremely important in the performance of maintenance, one must also consider the maintenance labour-hours expended in the process. Elapsed times can be reduced, in some instances by applying additional human resources in the accomplishment of specific tasks. However, this may turn out to be an expensive trade-off, particularly when high skill levels are required to perform tasks which result in less overall clock time. In other words, maintainability is concerned with the *ease* and *economy* in the performance of maintenance. As such, an objective is to obtain the proper balance among elapsed time, labour time, and personnel skills at a minimum maintenance cost. Thus, some additional measures must be employed. According to Blanchard *et al.* (1995) the following measured could be used:

1. Maintenance labour-hours per system operating hour (MLH/OH);
2. Maintenance labour-hours per cycle of system operation ($MLH/cycle$);
3. Maintenance labour-hours per month ($MLH/month$); and
4. Maintenance labour-hours per maintenance task (MLH/MT).

Any of these factors can be specified in terms of mean values. For example, $\overline{MLH}_c$ is the mean corrective maintenance labour-hours, expressed as (Hunt, 1993):

$$\overline{MLH}_c = \frac{\sum (\lambda_i)(MLH_i)}{\sum (\lambda_i)}$$

where: λ_i is the failure rate of the i th item (failures/h), and MLH_i is the average maintenance labour-hours necessary to complete repair of the i th item.

Additionally, the values for mean preventive maintenance labour-hours and mean total maintenance labour-hours (to include preventive and corrective maintenance) can be calculated on a similar basis. These values can be predicted for each echelon or level of maintenance and are employed in determining specific support requirements and associated cost.

21.3.3 Maintenance Frequency Factors

Based on the discussion thus far, it is obvious that reliability and maintainability are very closely related. The reliability factors, $MTBF$ and λ , are the basis for determining the frequency of corrective maintenance. Maintainability deals with the characteristics in system design pertaining to minimising the corrective maintenance requirements for the system when it assumes operational status later. Thus, in this area, reliability and maintainability requirements for a given system must be compatible and mutually supportive.

In addition to the corrective maintenance aspect of system support, maintainability also deals with the characteristics of design that minimise (if not eliminate) preventive maintenance requirements for that system. Sometimes, preventive maintenance requirements are added with the objective of improving system reliability (e.g., reducing failures by specifying selected component replacements at designated times). However, the introduction of preventive maintenance can turn out to be quite costly if not carefully controlled. Further, the accomplishment of too much preventive maintenance (particularly for complex systems/products) often has a degrading effect on system reliability as failures are frequently induced in the process. Hence, an objective of maintainability is to provide the proper balance between corrective maintenance and preventive maintenance at least overall cost. According to Blanchard and Lowery (1969) the most frequently used maintainability measure of this type are as follows.

Mean time between maintenance (MTBM). MTBM is the mean or average time between all maintenance actions (corrective and preventive) and can be calculated as

$$MTBM = \frac{1}{1/MTBM_u + 1/MTBM_s}$$

where $MTBM_u$ is the mean interval of unscheduled (corrective) maintenance and $MTBM_s$ is the mean interval of scheduled (preventive) maintenance. The reciprocals of $MTBM_u$ and $MTBM_s$ constitute the maintenance rates in terms of maintenance actions per hour of system operation. $MTBM_u$ should approximate $MTBF$, assuming that a combined failure rate is used which includes the consideration of primary inherent failures, dependent failures, manufacturing defects, operator and maintenance induced failures, and so on. The maintenance frequency factor, $MTBM$, is a major parameter in determining system achieved and operational availability.

Mean time between replacements (MTBR). MTBR, a factor of MTBM, refers to the mean time between item replacements and is a major parameter in determining spare-part requirements. On many occasions, corrective and preventive

maintenance actions are accomplished without generating the requirement for the replacement of a component part. In other instances, item replacements are required, which in turn necessitates the availability of a spare part and an inventory requirement. Additionally, higher levels of maintenance support (*i.e.*, intermediate and depot levels) may also be required.

In essence, *MTBR* is a significant factor, applicable in both corrective and preventive maintenance activities involving item replacement, and is a key parameter in determining logistic support requirements. A maintainability objective in system design is to maximise *MTBR* where feasible.

21.3.4 Maintenance Cost Factors

For many systems/products, maintenance cost constitutes a major segment of total life-cycle cost. Further, experience has indicated that maintenance costs are significantly affected by design decisions made throughout the early stages of system development. Thus, it is essential that total life-cycle cost be considered as a major design parameter beginning with the definition of system requirements.

Of particular interest in this chapter is the aspect of *economy* in the performance of maintenance actions. In other words, maintainability is directly concerned with the characteristics of system design that will ultimately result in the accomplishment of maintenance at minimum overall cost.

When considering maintenance cost, according to Blanchard and Lowery (1969), the following cost-related indices may be appropriate at criteria in system design:

1. Cost per maintenance action (\$/month);
2. Maintenance cost per system operating hour (\$/OH);
3. Maintenance cost per month (\$/month);
4. Maintenance cost per mission or mission segment (\$/mission); and
5. The ratio of maintenance cost to total life-cycle cost.

21.3.5 Related Maintenance Factors

It is evident from the analysis of maintenance process, performed earlier in the text, that there are a number of additional factors that are closely related to and highly dependent on the maintainability measures described. According to Blanchard these include various logistics factors, such as:

1. Supply responsiveness or the chance of having a spare part available when needed, supply lead times for given items, levels of inventory, and so on;
2. Test and support equipment effectiveness, which is the reliability and availability of test equipment, test equipment utilisation, system test thoroughness, and so on;
3. Maintenance facility availability and utilisation;
4. Transportation times between maintenance facilities; and
5. Maintenance organizational effectiveness and personnel efficiency.

There are numerous other logistics factors that should also be specified, measured, and controlled if the ultimate mission of the system is to be fulfilled. Many examples could be found where the interactions between the prime system and elements of support are critical, and both areas must be considered in the establishment of system requirements during conceptual design. Maintainability, as a characteristic in design, is closely related to the area of system support since the results of maintainability directly affect maintenance requirements. Thus, when specifying maintainability factors, one should also address the qualitative and quantitative requirements for system support in order to determine the effects of one area on another.

21.4 Empirical Data and Maintainability Measures

During the design, acquisition and operational phase of many engineering systems a large number of maintainability tests and predictions are conducted by maintainability engineers and managers in order to collect data relative to the length of time needed for the successful completion of the maintenance task considered. Thus, the final product of this effort is a set of numbers, denoted as dmt_i , where $i = 1, \dots, n$, each of which represents a length of time needed for the successful completion of the task analysed, when it is performed as specified. These data are the starting point for the statistical inference about one or more maintainability measures like:

- Maintainability function, $M(t)$;
- Percentual restoration time, TTR_p , $TTR_{95} = t = M_{\max}$; and
- Mean duration of maintenance task, $MDMT$.

The measures of maintainability addressed above provide very useful information for design, operation and maintenance engineers relative to the planning of logistic support resources (personnel, tools, equipment, facilities and similar), provision of which has a great impact on the logistic delay time and consequently to the operational availability of a product or system.

21.4.1 Possible Approaches to Analysis of Existing Data

Statistical inference is, generally speaking, a process of drawing conclusions about an entire population of similar objects, events or tasks, based on a sample of a few. Two following approaches to statistical inference are mainly used:

1. Parametric, which is primarily concerned with inference about certain summary measures of distributions (mean, variance and similar). This approach is based on explicit assumptions about the normality of population distributions and parameters.
2. Distribution, which is concerned with inference about entire probability distribution, free from the assumptions regarding the parameters of the population sampled.

Both approaches will be examined here and addressed from the maintainability-engineering point of view.

21.4.2 Parametric Approach to Maintainability Data

Following the main statistical principles regarding the parametric approach, which are based on the central limit theorem, in today's maintainability engineering practice, the numerical value of the Mean Duration of Maintenance Task, $MDMT^*$, of a particular sample size n , is computed according to the following expression (Blanchard and Lowery, 1969):

$$MDMT^* = \sum_{i=1}^n \frac{dmt_i}{n}$$

As the result obtained represents the mean value of this particular sample size, which has been selected at random, it is necessary to determine the interval within which the mean of the entire population lies. Thus, if one is prepared to accept the chance of being wrong, say 10% of the time, which corresponds to the 90% confidence limit, then the upper limit of the mean time to repair, $MDMT^u$ should be determined according to the following equation Blanchard (1969):

$$MDMT^u = MDMT^* + z\left(\sigma / \sqrt{n}\right)$$

where σ represents the standard deviation of the obtained empirical data, $(\sigma / \sqrt{n})$ is known as a standard error, and the value of z is selected from the table for the normal distribution based on the confidence level desired. This practically means that, say for $z = 1.28$, there is a 90% chance that the $MDMT$ of the entire population is less than the value obtained for $MDMT^u$.

21.4.3 Distribution Approach to Maintainability Data

Generally, the empirical data available captures much more information than the described parametric approach is able to disclose. In order to utilise fully the information contained in the existing maintainability data the application of the distribution approach to their analysis (Knezevic, 1985; MIL-STD-470B, 1989) should be applied. According to this approach the maintainability measures are expressed through the probability distribution of the duration of maintenance task, DMT , which is treated as a distribution-free random variable. Research conducted shows that the distribution type and their parameters have a significant influence on the maintainability measures. This method reveals great improvement in the effectiveness of the information extracted from the existing empirical data.

According to the distribution approach, the existing maintainability data are used as a basis for the selection of one of the theoretical probability distributions such as Weibull, normal, exponential, lognormal and similar, in order to model the maintenance task considered more accurately. This could be achieved by applying one of the following methods:

1. Graphical, where special probability papers are used as a tool for statistical inference;

2. Grapho-analytical, where the graphical method is supported with some analytical techniques in order to increase the accuracy of the statistical inference; or
3. Analytical, where the rigorous mathematical procedure is set up in order to provide a high level of accuracy of the inference process.

More details about each method can be found in Knezevic (1996). Regardless of the method used, the final result is the selection of the most suitable family of existing theoretical probability distributions for the modelling of the maintenance task considered, and the determination of the corresponding parameters that fully define the specific member of the family.

In order to demonstrate the procedure for the determination of the maintainability measures the example given earlier (replacement of the wheel) will be used.

21.4.3.1 Analysis of Experimental Results

Empirical data available are given in the second column in Table 21.12 together with the corresponding values of the maintainability function, $M'(t_i)$, where $i=1, \dots, 10$ after the data have been rearranged in ascending order. Numerical values for the $M'(t_i)$ were determined according to the expression for the median rank (because the total number of data is less than 50; Knezevic, 1993):

$$M'(t_i) = \frac{(i - 0.3)}{(n + 0.4)} \quad n = 10$$

Table 21.12. Empirical data for t_i and corresponding $M'(t_i)$

Student i	Time t_i	$M'(t_i)$
1	230	0.0673
2	259	0.1635
3	279	0.2596
4	286	0.3558
5	321	0.4519
6	332	0.5481
7	351	0.6442
8	365	0.7404
9	397	0.8365
10	442	0.9327

In order to determine the maintainability characteristics, the scale parameter A_m and shape parameter B_m need to be determined from the theoretical probability distribution that defines this data best. In order to achieve that, the

empirical data related to the time to change the wheel, t_i , and the cumulative distribution function, $M'(t_i)$, were plotted on the Weibull probability paper.

As the plotted empirical data form a straight line the conclusion that the Weibull distribution could be used to model the data obtained can be drawn. The particular member of family is defined by the parameters: $A_m=350, B_m=3.4$.

As the parameters of the Weibull distribution have been identified, it is possible to determine all measures of maintainability for the maintenance task observed as shown in Table 21.13.

1. *Maintainability function* ($M(t)$):

$$M(t) = 1 - \exp \left[- \left(\frac{t}{A} \right)^B \right], \text{ where } t \geq 0$$

Table 21.13. Maintainability function, $M(t)$, at different values of t

Time t_i (s)	$M(t)$
0	0
100	0.01403
200	0.13857
250	0.27279
300	0.44682
350	0.63212
400	0.79291
450	0.90464
500	0.96535
600	0.99807

2. *Percentage restoration time* TTR_p :

$$t = A \left[- \ln(1 - M(t)) \right]^{1/B}$$

Restoration time by which 10% of maintenance tasks are complete:

$$DMT_{10} = t \text{ for which } M(t) = 0.1$$

$$t = 350 \left[- \ln(0.9) \right]^{1/3.4} = 180.56 \text{ s}$$

Restoration time by which 90% of maintenance tasks are complete.

$$DMT_{90} = t \text{ for which } M(t) = 0.9$$

$$t = 350 \left[- \ln(0.1) \right]^{1/3.4} = 447.30 \text{ s}$$

3. *Mean duration of maintenance task* ($MDMT$):

$$MDMT = E(DMT) = A \times \Gamma \left(1 + \frac{1}{\beta} \right)$$

where Γ Gamma is the gamma function. From tabulated values therefore;

$$\Gamma \left(1 + \frac{1}{\beta} \right) = 0.898 \rightarrow MTTR = A \times 0.989 = 314.32 \text{ s}$$

Example 21.14: The empirical data given in Table 21.14 represents the length of time needed for the successful completion of a specific maintenance task, which is related to the two design alternatives, say A and B. The maintainability demonstration test was set up in such a way that 10, randomly selected, sufficiently trained mechanics were timed while performing a specific maintenance task regarding both alternatives, following the procedure given in the maintenance manual with full and free access to all support resources (tools, equipment, material, facilities and similar).

Table 21.14. Empirical maintainability data in minutes

I	1	2	3	4	5	6	7	8	9	10
ttr^{A_i}	206	167	323	193	128	181	218	249	151	275
ttr^{B_i}	186	92	273	158	35	121	221	360	64	486

21.4.3.2 Parametric Approach

According to this approach the mean duration of maintenance task, $MDMT^*$ and its upper limit, $MDMT^u$, with the confidence level of 85%, for both alternatives could be calculated by making use of the above equations and the values in Table 21.14.

Table 21.15. Maintainability measures extracted by the parametric approach

Alternative	$MDMT^*$	$MDMT^u$
A	200	214.8
B	200	245.7

The information calculated from the existing empirical data is almost all that can be extracted from this data. Making use of the results given in Table 21.15, some limitations of this approach are discussed below:

1. As both design alternatives have an identical value for $MDMT^*$, according to this maintainability measure, either alternative could be recommended for adoption;
2. Assuming that the contractual requirement was that $MDMT \leq$ say 50 minutes with a chosen confidence level of 85 %, based on the calculated values for $MDMT^u$, there is no a clear winner among competing alternatives, which practically means that both designs have equal legal right, despite the fact that alternative A has some advantages ($MDMT_u^A < MDMT_u^B$); and
3. The information obtained is not sufficient enough for the determining and plotting of the maintainability function for either alternative.

21.4.4 Distribution Approach

According to this approach, the empirical data available are used for the determination of the best-fit theoretical probability distribution (Weibull, normal,

exponential, and lognormal) to represent the maintenance task analysed. The results obtained, by using software “PROBCHAR” (Knezevic, 1993), are listed in Table 21.16.

Table 21.16. Maintainability measures extracted by distribution approach

	Alternative A	Alternative B
Distribution type	Normal	Weibull
Scale parameter	200	215
Shape parameter	50	1.25
<i>MDMT</i>	200	200
<i>DMT</i> ₁₀	135.9	35.5
<i>DMT</i> ₅₀	200	160.4
<i>DMT</i> ₉₀	264.1	419
Standard deviation	45.4	139.5

Comparing the data listed in Tables 21.15 and 21.16, it can be easily seen that the distribution method is able to extract much more information from empirical data, than the parametric method. One of its many advantages is that it is possible to determine the length of restoration time up to which 10, 50, 90, or any other percentage of maintenance task attempted will be successfully completed, regardless of the underlying theoretical distribution.

Maintainability function for both design alternatives, based on Equation 21.1 and the specific distributions and parameters selected (see Table 21.16) are defined as

$$M^A(t) = \int_{-\infty}^t \frac{1}{50\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-200}{50}\right)^2\right] = \Phi\left(\frac{t-200}{50}\right)$$

For alternative A, where Φ is a Laplace function for standardized normal variable, numerical values of which could be obtained from statistical tables, and for alternative B as:

$$M^B(t) = 1 - \exp\left(-\left(\frac{t}{215}\right)^{1.25}\right)$$

Clearly, the amount of information extracted from the existing empirical maintainability data in the latter case is much higher and potentially more beneficial to the decision maker, regarding maintenance management and logistic support issues. Also some of additional maintainability measures that could be extracted from the data, like restoration success, are providing a new light in maintainability studies.

21.5 Maintainability Engineering Predictions

“The only way to solve the problem would be to guess the outline, the shape, the quality of answer...We have no excuse that there are not enough experiments, it has

nothing to do with experiments.... We should not even have to look at experiments.... It is like looking in the back of the book for the answer.” Richard Feynman (Gleick, p. 1993, 303)

21.5.1 Introduction

Experience tells us that the biggest opportunity to make an impact on maintainability characteristics of any engineering system is at the design stage. Consequently, the biggest challenge for the maintainability engineers is to predict quickly and accurately the maintainability measures of the future maintenance task at the early stage of design, when changes and modifications are possible at almost no extra time and cost.

This is a very difficult engineering task due to multi-dimensional interactions between the sequences of activities within each task and the arrangements for the sharing of maintenance resources. Thus, the main objective of this text is to present a methodology for the fast and accurate engineering predictions of maintainability measures that could be used at the early stages of the design process for the maintenance tasks of the future systems based on the corresponding measures related to comprising maintenance activities.

The biggest challenge facing maintainability engineers is to predict maintainability measures related to maintenance tasks of:

- The future engineering systems at the early stage of design;
- The benefit of modifications on existing engineering systems.

In the text below, the new methodology for the fast and accurate prediction of maintainability measures and the identification of resources needed for the successful completion of maintenance tasks considered is presented. The proposed method is based on the maintainability measures related to the comprising maintenance activities, and the maintenance activities block diagram which is applicable to maintenance task whose consisting activities are performed: simultaneously, sequentially, and combined. The method presented could be successfully used at the very early stage of design when most of the information available is based on the previous experience, as well as, at the stage when design is completed and tests are performed in order to generate a maintainability data for the adopted configuration of the system.

21.5.2 Concept of the Maintainability Block Diagram

According to this methodology, each maintenance task is considered as a set of consisting maintenance activities. In order to analyze the maintainability characteristics of a maintenance task under consideration, the concept of Maintainability Block Diagram, MBD is introduced (Knezevic, 1994). This is a diagrammatical representation of the maintenance task where each of consisting maintenance activities is represented by a box. The relationship between boxes is determined by the order in which each of them has to be executed. The structure of an MBD for particular maintenance task is primarily inherited from design, although in some cases it could be altered by adopted maintenance policy. The time needed for the completion of each activity is irrelevant to the size of the box.

Based on the sequence in which maintenance activities are performed, according to Knezevic (1996), all maintenance tasks could be classified and defined as:

- 1. Simultaneous Maintenance Task represents a set of mutually independent maintenance activities, all of which are performed concurrently;
- 2. Sequential Maintenance Task represents a set of mutually dependent maintenance activities, all of which are performed in the predetermined order; and
- 3. Combined Maintenance Task represents a set of maintenance activities, some of which are performed in sequence and some simultaneously.

Regardless of the type of maintenance task the following symbols will be used here in order to derive the expressions for the prediction of maintainability characteristics: DMA_i , random variable which stands for the Duration of Maintenance Activity i , $MA_i(t)$, cumulative probability of completion of i th maintenance activity; nca is the number of consisting maintenance activities.

The methodology presented here provides a facility for the prediction of these characteristics, based on the corresponding characteristics of the consisting maintenance activities, in one hand, and the sequence of their execution, on the other. Thus, the time needed for the completion of each maintenance task is defined by the random variable DMT, and corresponding time related to each maintenance activity is defined by the random variable DMA

21.5.2.1 Simultaneous Maintenance Task

“Simultaneous maintenance task represents a set of mutually independent maintenance activities all of which are performed concurrently” (Knezevic, 1996).

The above definition fully describes the relationship between maintenance activities and clearly states that all activities are starting at the same instance of time and they are performed simultaneously and independently of each other. The maintenance task is completed then, and only then, when all individual activities have been successfully completed, as illustrated in Figure 21.4.

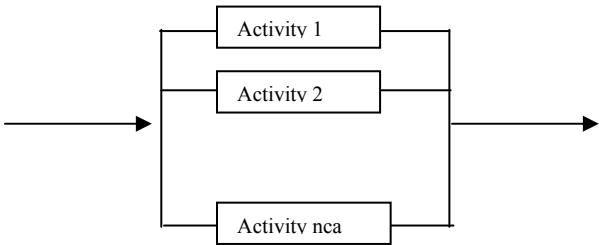


Figure 21.4. MBD for simultaneous maintenance task

In order to illustrate the simultaneous maintenance task a typical pit stop of a Formula 1 racing car will be analysed. During the course of a race, cars will make

one or more stops at their pits. Each stop consists of the preventive replacement of a set of four tyres, cleaning the driver's visor, cleaning the side pods and fuel refilling. As the race is only fought on the track and lost in the pit, all activities related to the pit stop are performed simultaneously. Speed of performing this task could mean, occasionally, the difference of a few places up at the end of the race.

21.5.2.2 Sequential Maintenance Task

"Sequential maintenance task represents a set of mutually dependent maintenance activities all of which are performed in the predetermined order." (Knezevic, 1996).

The above definition fully describes the relationship between maintenance activities and clearly states that each subsequent activity starts after the successful completion of the previous one. Thus, none of the subsequent activities can be originated and performed before the completion of the previous one. The maintenance task is completed then, and only then, when the last maintenance activity has been successfully completed, as shown in Figure 21.5.

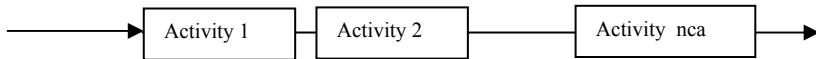


Figure 21.5. MBD for sequential maintenance task

A typical example of the sequential maintenance task is the replacement of the wheel of the car, where the activities have to be performed in the strictly determined sequence (see Section 21.3). Clearly, it is impossible to remove the wheel before the nuts/screws are undone and the wheel lifted off the ground.

21.5.2.3 Combined Maintenance Task

"Combined maintenance task represents a set of maintenance activities some of which are performed in sequence and some simultaneously." (Knezevic, 1996)

The above definition fully describes the relationship between maintenance activities within one task, and clearly states that activities are performed in combined order, as shown in Figure 21.6. Most maintenance tasks belong to this category, especially today when engineering systems are becoming more complex and consequently their maintenance tasks require higher levels of specialization.

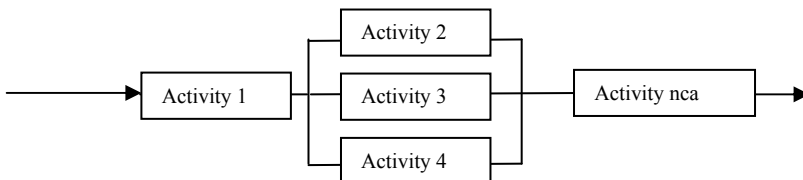


Figure 21.6. MBD combined maintenance task

A typical example of the combined maintenance task is the 6,000 (10,000) kilometers service to motor vehicles required by their producers. This task consists of activities that are related to the engine, transmission, brakes, body panels, electrical system and so forth. Thus, all of the required activities could be performed in predefined sequence that incorporates simultaneous and sequential execution of comprising maintenance activities.

21.5.3 Derivation of the Expression for the Maintainability Function

Having classified all maintenance tasks into three categories with respect to the timing of the execution of their comprising activities, it is necessary now to present a method for their prediction, at the early stage of a design process, when no mock-ups exist.

The definition of simultaneous task clearly states that all activities are starting at the same instant of time and they are performed simultaneously but independently of each other. The maintenance task is completed when all consisting activities have been completed.

Maintainability measures of the task whose consisting activities are performed simultaneously can be derived from the corresponding measures of consisting activities. Thus, the maintainability function of the maintenance task represents the probability that the task considered would be successfully completed by certain instance of time, t , $M(t) = P(DMT \leq t)$. At the same time, $M(t)$ could be represented as an intersection of events whose probabilities of occurrence are defined by the cumulative probabilities, $MA_i(t) = P(DMA_i \leq t)$. As in this case random variables DMA_i , where $i = 1, \dots, nca$, represent independent events, the maintainability function of the task $M(t)$, could predicted in accordance to the following expression:

$$\begin{aligned}
 M(t) &= P(DMT \leq t) \\
 &= P(DMA_1 \leq t \cap DMA_2 \leq t \cap DMA_3 \leq t \cap \dots \cap DMA_{nca} \leq t) \\
 &= P(DMA_1 \leq t) \times P(DMA_2 \leq t) \times \dots \times P(DMA_{nca} \leq t) \\
 &= \prod_{i=1}^{nca} P(DMA_i \leq t) \\
 &= \prod_{i=1}^{nca} MA_i(t)
 \end{aligned}$$

It is necessary to underline that the above expression could be very complex. The reason for this is the fact that the activity completion functions, $MA_i(t)$, where $i = 1, \dots, nca$, can be defined through any of the theoretical probability distributions (exponential, normal, lognormal, Weibull and similar) and in the majority of cases their product does not constitute any of the well known distribution functions.

Assuming that a maintenance task under consideration:

- Consists of four activities, A_1 , A_2 , A_3 and A_4 ;
- All four activities are performed simultaneously; and

- A maintainability functions of all activities are defined by the normal distribution with parameters $\mu_i=40$ min and $\sigma_i=15$.

It is possible to predict what, for example, the DMT_{90} of the future maintenance task will be. Hence, for the simultaneous task, the maintainability function can be derived as follows:

$$M(t) = P(DMT \leq t) = P(DMA_1 \leq t) \times P(DMA_2 \leq t) \times P(DMA_3 \leq t) \times P(DMA_4 \leq t)$$

For DMT_{90} , the individual activity probabilities must multiply to give 0.9 (because all four activities share the same maintainability function):

$$0.9 = M(t) = P(DMA_1 \leq t) \times P(DMA_2 \leq t) \times P(DMA_3 \leq t) \times P(DMA_4 \leq t)$$

Each individual probability is thus, $\sqrt[4]{0.9} = 0.974$. In order to find the time (t) by which 97.4% of maintenance activities, of a certain type will be successfully accomplished, the standardized normal variable is used, which fixed the value of $z = 1.9442$. Thus the DMT_{90} for the maintenance task is

$$1.9442 = \frac{t - 40}{15} \rightarrow t = DMT_{90} = 69.16 \text{ min}$$

All other maintainability measures of this task could be predicted in similar manner, using the method proposed.

21.5.3.1 Maintainability Measures for the Sequential Maintenance Task

The definition for the sequential task fully describes the relationship between comprising maintenance activities and clearly states that each subsequent activity starts after the completion of the previous one. Thus, none of the subsequent activities can be performed before the successful completion of the previous one. The maintenance task is completed when the last activity has been completed.

Maintainability measures of a maintenance task, whose consisting activities are performed sequentially, from the point of view of maintainability, can be derived from the maintainability measures of its comprising activities. Thus, the maintainability function of the maintenance task, $M(t)$, whose consisting activities are performed in a predetermined sequence represents the probability that the maintenance task under consideration will be completed within interval of time, from zero to t inclusive, could also be represented as a sum of a sequential independent random variables DMA_i where $i = 1, \dots, nca$. Consequently, the maintainability function of the task is equal to the nca th convolution of comprising activities, thus

$$\begin{aligned} M(t) &= P(DMT \leq t) \\ &= P(DMA_1 + DMA_2 + DMA_3 + \dots + DMA_{nca} \leq t) \\ &= MA^{nca}(t) \end{aligned}$$

where $MA_i(t)$ represents the i th convolution of comprising maintenance activities. It could be determined in accordance to the following expression:

$$MA^i = \int_0^t MA^{i-1}(x) dMA_i(t-x) \quad \text{where } i = 1, \dots, nca$$

It is necessary to point out that the above expression is not a single equation; it is a simultaneous set of large a number of convolution integrals (Cox, 1962). All information about the properties of the maintenance task and its consisting activities including their number, probability distributions, their sequence and their DMT and many similar characteristics are stored in it. It represents a cumulative probability that a maintenance task initiated at time zero with activity 1 will be completed by time t , with the successful completion of the last activity.

It is necessary to underline that the above expression could be very complex, as it is a convolution of a large number of contributing activities, as shown below:

$$MA^1(t) = MA_1(t)$$

$$MA^2(t) = \int_0^t MA^1(x) dMA_2(t-x)$$

$$MA^3(t) = \int_0^t MA^2(x) dMA_3(t-x)$$

...

$$MA^n(t) = \int_0^t MA^{n-1}(x) dMA_n(t-x)$$

Further reason for the complexity of the above expressions is the fact that the maintainability functions of comprising activities, $MA_i(t)$ where $i=1, nca$, can be defined by any of the theoretical probability distributions. Consequently, in the majority of cases the final convolution function, for the task, cannot be calculated analytically with ease.

Assuming that a maintenance task under consideration consists of four sequentially performed activities, A_1, A_2, A_3 and A_4 , each defined by the normal distribution with parameters $\mu_i=40$ min and $\sigma_i=15$, it is possible to predict, for example, the value of DMT90 of the future maintenance task.

In this particular case each of four maintenance activities are defined by an identical probability function, thus (Knezevic, 1996):

$$MA_1(t) = MA_2(t) = MA_3(t) = MA_4(t) = \Phi\left(\frac{t-40}{15}\right)$$

For the normal probability distribution, the maintainability function of the task is derived using the following expression:

$$M(t) = \Phi\left(\frac{t - \mu_{task}}{\sigma_{task}}\right)$$

The parameters A and B for the overall task maintainability function are calculated

as follows: $\mu_{task} = \sum_{i=1}^{nca} \mu_{activity_i} = 40 + 40 + 40 + 40 = 160$

$$M(t) = \Phi\left(\frac{t - \mu_{task}}{\sigma_{task}}\right) \text{ and } \sigma_{task} = \sqrt{15^2 + 15^2 + 15^2 + 15^2} = 30.$$

Thus, $M(t) = \Phi\left(\frac{t-160}{30}\right)$, and for our example $0.9 = \Phi\left(\frac{t-160}{30}\right)$

From the standardised normal distribution tables, the value for $z = 1.2819$ thus

$$1.2819 = \frac{t - 160}{30} \rightarrow DMT_{90} = 198.45 \text{ min.}$$

This means that the maintenance task considered, in 90% of trials will be successfully completed within 198.45 min.

Another measures of maintainability, like MDMT for example, could be also predicted by applying the method proposed.

21.5.3.2 Maintainability Measures for Combined Maintenance Task

As the definition for combined maintenance task suggests it is a combination of maintenance activities, some of which are performed simultaneously, and some of them following a predetermined sequence. Thus, the maintainability function for the combined maintenance task depends on the maintenance activities block diagram and the activity completion functions under consideration.

Most maintenance tasks belonging to this category, especially today when the engineering systems have become more complex and consequently require higher levels of specialization. In these cases a predefined expression for a generic maintenance task does not exist. For each single task it is necessary to build a MBD, and then derive the analytical expression for it.

To illustrate the method presented, the following hypothetical example will be used: maintenance task under consideration consists of nine identifiable maintenance activities, defined probabilistically with a data, in minutes, given in the Table 21.17.

Table 21.17. Maintainability measures for the combined task analyzed

Activity	Distribution	Scale	Shape	MDMA	DMA ₅₀	DMA ₉₀
1	Weibull	30	2.5	26.62	25.91	41.88
2	Weibull	10	5	9.18	9.29	11.82
3	Weibull	60	3.7	54.15	54.34	75.17
4	Weibull	30	4.1	27.23	27.43	36.77
5	Weibull	75	3.5	67.48	67.54	95.18
6	Weibull	50	4.1	45.38	45.72	61.28
7	Weibull	60	5.6	55.45	56.20	69.64
8	Weibull	45	3.2	40.30	40.13	58.40
9	Weibull	20	2.7	17.79	17.46	27.24

Maintainability Block Diagram for the task analysed is presented in Figure 21.7.

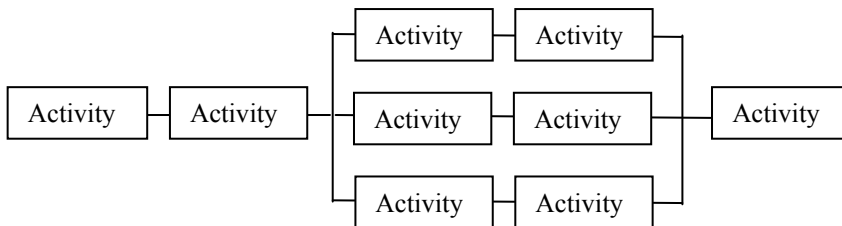


Figure 21.7. MBD for the maintenance task analysed

Making us of expressions for sequential and simultaneous maintenance task and the MBD, the expression for the combined maintenance task has been generated, thus:

$$\begin{aligned}
 M(t) &= P(DMT \leq t) \\
 &= P(DMA_1 + DMA_2 + \\
 &\quad \{[(DMA_3 + DMA_4] \cap [(DMA_5 + DMA_6] \cap [(DMA_7 + DMA_8]\} \\
 &\quad + DMA_9 \leq t) \\
 &= P(DMA^{1+2} + \{[DMA^{3+4}] \times [DMA^{5+6}] \times [DMA^{7+8}]\} \\
 &\quad + DMA_9 \leq t)
 \end{aligned}$$

Individual elements of the above expression could be found in the following way:

$$\begin{aligned}
 MA^{1+2}(t) &= P(DMA_1 + DMA_2 \leq t) = \int_0^t MA_1(x) dMA_2(t-x) \\
 MA^{3+4}(t) &= P(DMA_3 + DMA_4 \leq t) = \int_0^t MA_3(x) dMA_4(t-x) \\
 MA^{5+6}(t) &= P(DMA_5 + DMA_6 \leq t) = \int_0^t MA_5(x) dMA_6(t-x) \\
 MA^{7+8}(t) &= P(DMA_7 + DMA_8 \leq t) = \int_0^t MA_7(x) dMA_8(t-x)
 \end{aligned}$$

Although the above expressions are uniquely defined, the difficulties stem from the multidimensional integration in time domain over all consisting maintenance activities. The Monte Carlo method is specially oriented towards the evaluation of multidimensional averages on complex situations (Dubí, 2003). Hence, it provides a very useful computational tool for constructing and analysing complex MBD. The maintainability function for the example analysed has been obtained through the application of the Monte Carlo simulation method, and it is shown in Figure 21.8.

In summary this example shows how a complex and real maintainability engineering problems could be dealt with by applying the method proposed.

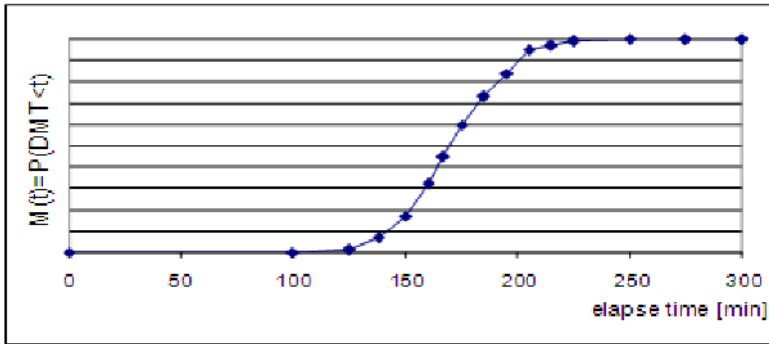


Figure 21.8. Maintainability function for the combined maintenance task

21.5.4 Maintainability Characteristics for Different Design Options

In order to demonstrate the applicability of the method proposed an example related to the maintenance task of changing a wheel on a small passenger car will be used. The list of specified activities that need to be completed and their sequence are shown in Table 21.8.

According to the experience gained from the previous model of the car under consideration, and the layout of the present design, the predicted values for the mean time to complete each of 11 consisting activities, $MDMA_i$, could be generated. At this stage of design the exact type of the probability distribution that could be used to represent each activity is not known. Hence, it is not unreasonable to assume that all maintenance activities could be modelled by the normal distribution. Numerical values for the standard deviation, $SDDMA_i$ for each task, are reflecting the spread of data among all potential users, their physical and mental differences, as well as the influence of climate, solar radiation, rain, sun, and many other factors that might make impact on it. The experience-based values, which take into consideration the variability of the factors that define the environment under which the task is performed, are given in the Table 21.18.

Table 21.18. Predicted values for consisting activities in seconds

Activity	$MDMA_i$	$SDDMA_i$
1	45	15
2	15	3
3	60	20
4	10	3
5	20	7
6	20	7
7	60	20
8	10	3
9	20	7
10	10	3
11	60	20
Task	330	40

The main objective of this exercise is to derive the expressions for the maintainability measures of the task analyzed based on predicted activity completion functions of the $MA_I(t)$, where $I = 1, \dots, 11$.

Based on the task description stated above, and the types of maintenance tasks given above, it is not difficult to conclude that the task considered is the sequential maintenance task. Consequently, its maintainability function could be obtained by applying equation for the sequential maintenance task. Thus, in this particular example it will have the following form:

$$\begin{aligned}
 M(t) &= MA^{11}(t) \\
 &= \int_0^t \frac{1}{B^{11} \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{t - MDMA^{11}}{SDDMA^{11}} \right)^2 \right] dt \\
 &= \int_0^t \frac{1}{42 \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{t - 330}{40} \right)^2 \right] dt \\
 &= \Phi \left(\frac{t - 330}{40} \right)
 \end{aligned}$$

where $MDMA^{11} = \sum_{i=1}^{11} MDMA_i$, and $SDDMA^{11} = \sqrt{\sum_{i=1}^{11} SDDMA_i^2}$

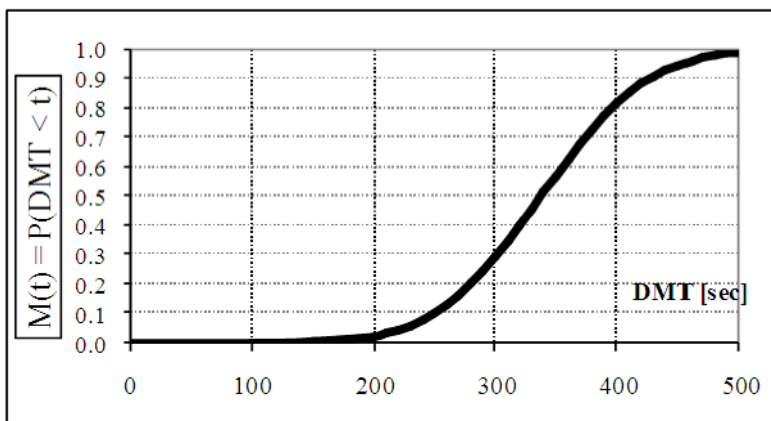


Figure 21.9. Predicted maintainability function design-0

Once the expression for the maintainability function has been obtained all other maintainability measures could be determined according to expressions already given.

In order to compare the values obtained from the real life test and the predicted values obtained by the proposed methodology, the numerical values for the several maintainability measures are given in Table 21.19.

Table 21.19. Predicted maintainability measures in seconds for the task analyzed

MDMT	DMT ₁₀	DMT ₅₀	DMT ₉₀
330	275.5	320	379.5

The similarity between results obtained in Example 21.14 and the data shown in Table 21.19 clearly illustrates the accuracy and the usefulness of the methodology proposed for the prediction of the maintainability measures of related maintenance tasks for the future systems at a very early stage of design. It is necessary to stress that the predicted values are obtained within several seconds of calculation without any additional cost.

The second major advantage of the methodology proposed is the possibility for the quantitative evaluation of the different design options to the maintainability measures. The results for the basic design related to $M(t)$ are shown in Figure 21.9.

In the above example it is possible to examine quantitatively the changes in the maintainability characteristics by making some of the following design changes:

Alternative 1: place a spare wheel in engine department. This will certainly influence the activities 1 and 11, because it would not be necessary to remove the content of the boot and remove the shelf in order to access the spare wheel. Let us assume that the new configuration will reduce the $MDMA_1$ to 20 s, and $MDMA_{11}=15$. The consequences of this design change to $M(t)$ are shown in Figure 21.10 and the predicted values for the activities are given in Table 21.10.

Table 21.20. Predicted values for consisting activities in seconds

Activity	MDMA _i	STDEVMA _i
1	20	7
2	15	3
3	60	20
4	10	3
5	20	7
6	20	7
7	60	20
8	10	3
9	20	7
10	10	3
11	15	5
TASK	260	32.5

$$M(t) = \int_0^t \frac{1}{32.5\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{t-260}{32.5} \right)^2 \right] dt = \Phi \left(\frac{t-260}{32.5} \right)$$

Alternative 2: Have a wheel attached to the hub by one central wheel nut. This should decrease the lengths of activities 3, 6, 7, and 9 because there will be only one nut instead of four bolts. If we assume that the new configuration will reduce the MDMA₃ to 30 s, MDMA₆ from 20 s to 10, MDMA₇ to 30 s and finally MDMA₉ to 9 s. The predicted values of MDMT and SDTT for the new configuration are given in Table 21.21 and the consequences of this design change to M(t) are shown in Figure 21.11.

$$M(t) = \int_0^t \frac{1}{30.5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-239}{30.5}\right)^2\right] dt = \Phi\left(\frac{t-239}{30.5}\right)$$

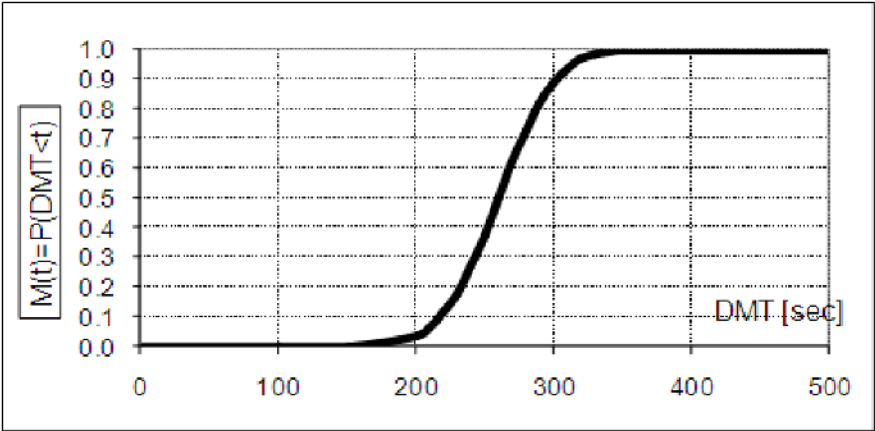


Figure 21.10. Predicted maintainability function design – 1

Table 21.21. Predicted values for consisting activities in seconds

Activity	MDMA _i	SDDMA _i
1	45	15
2	15	5
3	30	20
4	10	3
5	20	7
6	10	3
7	30	10
8	10	3
9	20	7
10	9	3
11	50	20
TASK	270	30.5

$$M(t) = \int_0^t \frac{1}{30.5\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{t - 270}{30.5} \right)^2 \right] dt = \Phi \left(\frac{t - 270}{30.5} \right)$$

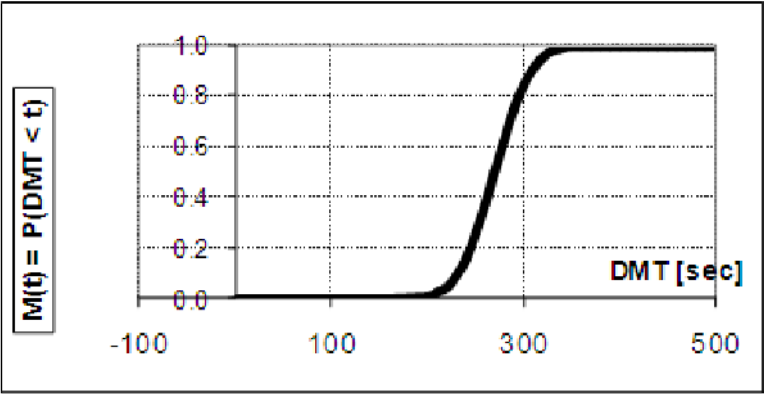


Figure 21.11. Predicted maintainability function design – 2

Alternative 3: Keep original configuration, but use a hydraulic jack instead of a mechanical one. This change will affect activities 5 and 8 in the following way: $MDMA_5=10$, $MDMA_8=3$. The new values for the mean time to replace the wheel and the standard deviation are given in the penultimate column of Table 21.22 and the consequences of this design change to $M(t)$ are shown in Figure 21.12.

Table 21.22. Predicted values for consisting activities in seconds

Activity	$MDMA_i$	$SDDMA_i$
1	45	15
2	15	5
3	60	20
4	10	3
5	10	3
6	20	7
7	60	20
8	3	1
9	20	7
10	10	3
11	50	20
TASK	303	39.5

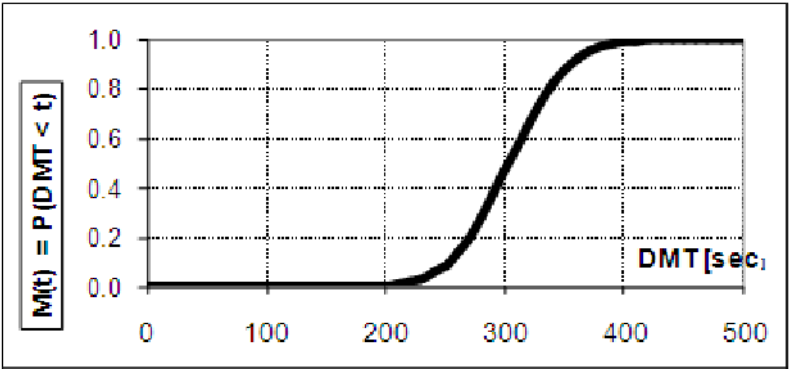


Figure 21.12. Predicted maintainability design – 3

Alternative 4: Combined possibilities 1 and 3.
The impact of these changes to the maintainability function is shown graphically in Figure 21.13 and the predicted values of the activities are given in Table 21.23.

Table 21.23. Predicted values for consisting activities in seconds

Activity	MDMA _i	SDDMA _i
1	20	6
2	15	3
3	60	20
4	10	3
5	10	3
6	20	7
7	60	20
8	3	1
9	20	7
10	10	3
11	15	5
TASK	243	31.6

$$M(t) = \int_0^t \frac{1}{31.6\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{t - 243}{31.6} \right)^2 \right] dt = \Phi \left(\frac{t - 243}{30.5} \right)$$

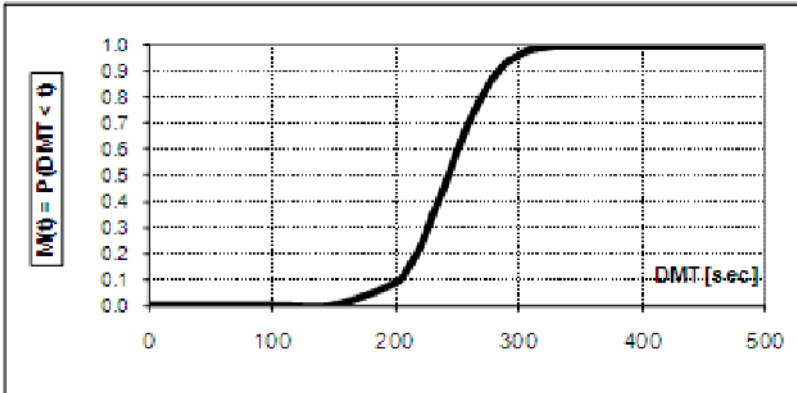


Figure 21.13. Predicted maintainability function design – 4

Table 21.24. Derived distribution parameters for possibilities examined

	Engineering design alternatives				
Maintainability measures	0	1	2	3	4
MDMT	330	260	280	313	243
SDDMT	40	32.5	30.5	39.5	31.6
DMT_{10}	275.5	215.8	239.4	259.8	200.5
DMT_{50}	330	260	280	313	243
DMT_{90}	379.5	299.2	316.4	361.2	282
DMT_{95}	392.7	312.7	327.2	375.4	292.8

In Table 21.24, 0 stands for the base-line design, and 1, 2, 3 and 4 for the four possible design changes

Clearly, the proposed methodology enables a design team to quantify quickly the consequences of the design solution chosen as well as the consequences of the possible design changes to the maintainability measures of the maintenance task under consideration at a very early stage when the changes could be easily implemented at almost no extra cost.

21.6 Maintainability Engineering Management

There has always been a Chief Pilot on every Boeing model, but 777 is the first Boeing model with a Chief Mechanic. This certainly illustrates the recognition of the importance of the maintenance process to successful airline operation³.

With the increase of the rate and pace of technological advances, more and more individual companies and whole industries have been and are being forced to abandon complacency and to look into the future. Action, in such cases, usually starts with increased emphasis on: research, development and design.

The importance of the maintainability engineering management function, MEMF, in the design process in a company is directly proportional to the importance of the design function. A strong, competent design function is essential in areas of rapid technological advances such as aerospace and weapon systems. The design function is also of great importance to producers of consumer goods (such as motor vehicles, office equipment, and household appliances), to machine-tool producers, and in many other similar areas. Design organizations are usually strong staff organizations in companies producing these systems. Although of lesser importance in companies producing simple systems or systems of stable, proven design, the design function is an important one in all production industries.

The design function within a company has certain responsibilities to the organization management. Working in the assigned system areas, the design function must create designs that are functional, reliable, maintainable, producible, timely and competitive.

Whenever possible, the designers are expected to use proven design techniques. When using the proven and familiar design methods cannot meet design objectives, the designers are expected to adapt their methods, introduce design practices from other industries, or use some of the new state-of-the-art materials and processes that are available. Since designers are usually, by disposition, creative, it is often difficult for them to resist trying something new, even though a proven technique is available. Designers are well known for their receptiveness to the efforts of parts and package suppliers' sales engineers, who are selling the outstanding merits of their new system. A major responsibility of the design management is the establishment of a system that makes it appreciably easier for designers to use proven designs than to use unproven designs.

Since it is often impossible to meet all design objectives to the maximum desired extent, designers are frequently required to trade off between them. By requiring unusually tight tolerances or by specifying an exotic material, they may improve reliability at the expense of maintainability. By not fully testing the ability of the design to function under the worst combinations of operational environment, the designers may take a chance on worsening maintainability so that the design disclosure can be released on schedule. Some of these compromises and trade-offs are unavoidable, and it is the design function that has both the information and responsibility to make the required decisions in these cases. Thus, the fact that

³ Source: Airliner, January–March, 1995

trade-offs have to take place makes the interaction between all participants within the design function essential, which practically means that final decisions should not be made without full support obtained by maintainability engineers, as a part of an integrated team.

21.6.1 Role of the Maintainability Engineering Management Function

In order for a design to provide an acceptable inherent maintainability, provisions must be made within the design concept and must continue throughout development to its completion.

However, while performing the maintainability analysis of the design proposed, maintainability personnel may discover and require correction of design features, errors, or omissions which affect feasibility. However, this is not a primary purpose of maintainability analysis. Maintainability is concerned with all design issues that are able to assure that a functionality of a feasible design could be easily, safely and economically maintained during operation under specified environments and other operating conditions.

The maintainability engineering management function works with the design function in several ways to achieve its objective. Maintainability acts at various times as a helper, and as a conscience of a design team, but very rarely should it act as an inspector (although it is still practice in many organizations).

As a part of an integrated design team, maintainability engineers perform certain analytical and statistical analyses. These include collection, analysis, and feedback of data on development hardware/software. Generally speaking, MEMF assists the design team in predicting and measuring inherent maintainability during the various design stages.

Maintainability serves as a conscience to the design function by closely participating in the design progress, in a concurrent manner, towards the specified maintainability goals. In addition, all trade-offs affecting maintainability are very closely examined.

Inevitably, MEMF assesses the design output in order to verify the proposed solution before the design effort can progress. Some of the checkpoints for such enforcement of maintainability requirements are the maintainability approval of solutions proposed by the designers, maintainability approval of actual usage of the approved maintenance resources, maintainability approval of design reviews and, finally, maintainability signature approval of the design-disclosure documents (drawings, specifications, and procedures).

The maintainability engineering management function may also include coordination of the design test programs, direction of independent maintainability test programs and actual conduct of all testing, identification and establishment of control systems for portions of the design which have special limitations, preparation of maintainability specifications applicable to suppliers, and the imposition of maintainability requirements on suppliers through review and approval of procurement documents.

21.6.2 MEMF Opportunities

Generally speaking, the more difficult the design assignment, the larger the maintainability effort is required on one hand, but larger impact on design could be made on the other.

The design problems encountered in an aircraft, a major weapon system, or a complex worldwide communications network require substantial design and maintainability efforts. At the same time if the design is well within the existing technology, simple, and has ample space, weight, and design-time allowances, then a relatively small maintainability effort may be adequate.

Thus, a major maintainability effort (major in the sense of being a substantial fraction of the design effort) is required under the following circumstances:

1. In the design of any very complex system; and
2. In the design of systems with very high maintainability requirements, particularly when the designers are working within severe space and weight limitations (submarines, spacecrafts, racing cars and so forth).

The general rule could be made that the more constraints on the design team and the tighter the constraints are, the greater is the required maintainability program. One of the major reasons for this relationship is that, under the pressure of these constraints, the design teams may intentionally or unintentionally neglect the maintainability requirements. A strong maintainability effort is needed both to assist and to check on the design teams in matters affecting maintainability requirements and measures.

During the design development the design teams must work out compromises between various requirements. The penalties for design teams not fully meeting required performance, schedule, cost, producibility, and other goals are much more immediate and certain than are the penalties for not fully meeting maintainability goals. Consequently, an MEMF must be a part of a team in order to call immediate and forceful attention to any deficiencies in the design provisions for maintainability. The presence of MEMF should assure that full consideration is given to maintainability requirements. Clearly, in some cases it may still be necessary to trade off a maintainability requirement for performance or for any other design characteristic, but such a trade off must be made with full knowledge of all possible consequences.

21.6.3 MEMF Obstacles

Very few integrated design teams deliberately skimp on provisions for attaining the full maintainability requirements in their creations. However, some of the following dangers are always present:

- Oversight;
- Lack of specific knowledge; and
- Rationalisation.

Each of these are briefly addressed below.

21.6.3.1 Oversight

This type of obstacle occurs in cases when the design teams fail to take care of one of those innumerable details that make up the completed design.

For example, the design teams are fully aware that a special fastener is required for a specific item but fails to indicate it on their drawing. Thus, if this oversight is not caught, a substantial delay in completion of a corresponding maintenance task might occur.

Further, it is very easy to fail to notice that the execution of maintenance tasks will, occasionally, be required to be performed under non-typical conditions like low/high temperature, chemical contamination, and similar when the completion of the tasks is significantly more difficult and often impossible.

21.6.3.2 Lack of Specific Knowledge

It is fair to say that members of design teams cannot know all there is to be known about everything connected with every design, nor do they have the time to verify every detail. Consequently, all design teams do what they can to check where they believe it necessary, and call in the experts in certain highly specialized areas such as use of specific maintenance resources, ergonomics, safety issues and similar.

For instance, the design teams may specify the use of specific test equipment or a tool which was the best available for the purpose the last time when there was a need for it. However, a new technology that functions better and that performs as well in excess of the design requirements may have become available.

21.6.3.3 Rationalisation

The design teams are usually pressed for time; hence on many occasions there is a honest belief that the design proposed will meet all of the design requirements including the maintainability requirements, but an additional series of tests should be run to be absolutely certain. Waiting for the test results will put the design behind schedule. It is easy for the design teams to work themselves into a frame of mind that the tests are not really necessary. This same practice of rationalisation includes explaining away test failures with such comments as "It was a testing error" or "It was an early design" or "The real environments will never be nearly that rigorous anyway."

When the consequences of not achieving the maintainability requirements are extremely dangerous, expensive in revenue terms, or jeopardize safety, reputation, or national security, then maintainability considerations become extremely important. When maintainability characteristics are of high priority, it cannot be left to good intentions or to chance. Thus, there must be an independent check and balance of every maintenance task (including those operations which the maintainability-quality function itself may perform, such as the writing of test procedures) and continuous attention paid to details. Hence, no organization and no person can be considered so good or so omnipotent that its or their output need not receive a searching independent analysis.

21.6.4 Design Methods for Attaining Maintainability

Those specific practices design teams may follow to achieve maintainability, which would not necessarily be followed if the main concern is in only getting the design to function, are discussed here. Thus:

- Accessibility;
- Modularity;
- Simplicity;
- Standardisation;
- Fool-proofing; and
- Inspectability.

Each of the above listed practices are briefly addressed below.

21.6.4.1 Accessibility

All equipment and sub-assemblies that require regular inspection should be located in such a way that they can be accessed conveniently and are fitted with parts that can be connected rapidly for all mechanical, air, electric and electronic connections.

For example, in the case of the TGV train, the roof panels can be rapidly dismantled, and lateral access panels and numerous inspection points allow for all types of progressive inspections and components in a short space of time. The auxiliary equipment in power cars and passenger cars is located so that work positions for maintenance staff are ergonomic and especially in such a way that several specialists can carry out maintenance work simultaneously without disturbing other maintenance staff at work.

Generally speaking, whenever possible, it should not be necessary to remove other items to gain access to items requiring maintenance especially high replacement items. Also, items like lubricates should be possible to replace or topped-up without disassembly.

21.6.4.2 Modularity

A modular approach is a fundamental guarantee of ease of replacement. Furthermore, this can only be achieved if interface equipment is standard. The range of physical orders of magnitude at input and output of each module ensure that no readjustments are required when they are incorporated in a unit of equipment. For example, SAAB Gripen's RM12 engine is modular in design which makes it easy and quick to inspect and replace only the necessary module.

21.6.4.3 Simplicity

Generally speaking, the simpler the design the better the maintainability properties. For example, a reduction in the quantity of parts or in the number of different parts used is a standard approach in trying to improve the maintainability.

For instance, no tools at all are required to open and close the service panels on the SAAB Gripen aircraft. All control lights and switches needed during the turnaround time are positioned in the same area, together with connections for the

communication with the pilot and refueling. On the same aircraft, the simple portable mini-hoist, is used for loading the external stores as well as for the engine replacement.

21.6.4.4 Standardisation

Standard fasteners, connectors, tools, and test equipment have usually been thoroughly tested and are less likely to cause problems. Consequently, designers should use standard parts as far as possible, *e.g.*, seals, nuts and bolts, especially high replacement rate items.

Since typical design teams are by nature creative people, it requires a great deal of restraint to stay with simple designs and continuous use of standard parts, tools and facilities.

21.6.4.5 Fool-proofing

Items that appear to be similar but are not usable in more than one application should be designed to prevent fitting to the wrong assembly. Incorrect assembly should be immediately obvious, not at a later stage such as the fitting of cover plates or during testing. Some of the following considerations should be inbuilt during design:

- If an item is secured with three or more equidistant fasteners, stagger their spacing;
- Ensure that shafts that are not symmetrical about all axes cannot be fitted wrongly, either end to end or rotationally;
- Where shafts of similar lengths are used, ensure that they cannot be interchanged *e.g.*, vary their diameters;
- Avoid using two or more pipe-fittings close together with the same end diameters and fittings;
- Where pipes are in close proximity to one another, ensure that the run of each pipe relative to the others is easily discernible;
- Flat plates should have their top and bottom faces marked if they need to be installed with a particular face upwards; and
- Springs of different rates or lengths within one unit should also have different diameters.

Thus, design teams should make it as difficult as possible to assemble or use their design incorrectly. When possible, cable lengths should be such that only the correct cable can reach the black-box connector. When the design is such that more than one cable can reach a black-box connector, the cable connectors should be of different sizes so that only the connector of the correct cable will fit. When a functional package is to be a space, full maintainability consideration must be given to the problems of removal and replacement by ordinary people under field conditions. If the design is such that it is extremely difficult to remove and replace an LRU/SRU, the probability that the maintenance task will not be completed successfully becomes substantial. If the design is such that it is possible to drop an attaching bolt, nut, or screw into a vital or inaccessible area, the probability that this will happen, given a number of opportunities, becomes high.

21.6.4.6 *Inspectability*

Whenever possible, designers should create a design that can be subjected to full non-destructive, functional checkout. For example, a circuit breaker can be functionally checked, whereas such a check on a fuse is destructive. Here the testing advantage must be weighed against the maintainability of the fuse. The ability to inspect important dimensions, joints, seals, surface finishes, and other non-functional attributes up to and past the assembly point where they are likely to be degraded is also a very important characteristic of a maintainable design.

Example 21.15: Integration between maintainability engineers with design teams existed from the start of the TGV project (French designed and built high speed train). A multi disciplinary design team was formed where maintainability engineers played an important and officially acknowledged role. They worked directly with rolling stock design engineers and provided them with the benefit of their experience thereby avoiding conflicts and delays. As a result, provisions and specifications for maintainability were built into technical documents defining rolling stock. They were based on systematic analyses of past experience and records of all technical provisions hindering maintenance procedures, tried and proven solutions which should be incorporated in new rolling stock.

However, the SNCF Rolling Stock Department undertook to integrate maintainability to an even greater degree. From the very start of research and for several years, maintenance specialists were deliberately assigned to work with rolling stock research, development and design departments. They ensured that the blueprint or CAD stage, solutions familiar to maintenance engineers were incorporated in practice in design specifications. At the same time they received and passed on to maintenance-departments the outcome of initial research work and diagrams which would be useful for developing maintenance procedures and staff training programs as well as for defining and setting up installations and equipment required for maintenance.

In this example some of the achievements made by the integrated design team are addressed below:

- Wheels are not subjected to a large degree of wear because wheel materials, geometry, lubrication and the limited forces to which they are subjected, especially with modern brake gear, are such that retro filing has been pushed back to beyond 450,000 km (roughly 280,000 miles). Even then, wheels are re-profiled simply and at low cost with numerically-controlled pit-mounted wheel lathes.
- Electric commutator motors which required monitoring, maintenance and replacement when they reached their overhaul period have now been replaced by more powerful self-commutated synchronous drive motors for the Atlantic TGV and asynchronous motors for all auxiliaries.
- Visual and instrument monitoring of commutators, brush replacement machining, replacement of worn or damaged commutators and especially rewinding of sections have all been eliminated. Monitoring of the mechanical part for the new motors is expected to be very simple and reliability of the associated static convertors fully under control.

- Destination indicator panels on the Southeast TGV train are manually-controlled mechanical systems: these have been replaced on the Atlantic TGV by micro-processor-based static liquid crystal displays, which are remote-controlled by radio.
- For the TGV, mechanised cleaning of the external body shell might have been problematic because of the contours of trainset ends. Maintainability was achieved by special kinematics built in to an automatic train washing machines and by an automatic mechanism controlling the rate of advance of trainsets.
- The innovation of retention toilets has not been a hindrance for operations nor for maintenance: underground automatic evacuation installations have been trouble-free.
- In order to protect components from dirt accumulation, electronic equipment and power circuits on the Atlantic TGV are cooled in sealed units filled with a liquid refrigerant and are thereby protected from direct contact with ventilation air to eliminate pollution and the difficult task of cleaning components.

Failure protection measures built into design:

- Built-in redundancy: much of the equipment on Paris-Southeast trainsets features built-in redundancy *e.g.*, in the power system in command and control circuitry and in technical and passenger comfort auxiliary equipment. Back-up components take over automatically if a failure occurs without causing any disruptions to train operation. This aspect of maintainability is enhanced on the Atlantic TGV by a facility which stores records of switchovers to back-up components and for some functions, data is transmitted to the maintenance center.
- Automatic monitoring: on all TGV trains, the main safety functions (fire detection, mechanical damage, instability) are monitored automatically by the train-borne system to safeguard against exceptional catalytic failures and prevent frequent costly checks that are unlikely to yield defects. This equipment for which fortunately there is only a very minute probability that it might be needed has its own automatic built-in test facility. To monitor the temperature of roller-bearing axle boxes on the high-speed line, the solution of automatic monitoring at 40-km (25-mile) intervals along the line has proved to be highly appropriate and will be used again on the Atlantic TGV line. Although the term 'hot box detector' has been used, in fact the system is an automatic infrared thermometric network for which data is computerized and processed centrally. In addition to its role as an emergency hotbox detector, the system can supply highly useful preventive data on any abnormal changes in axle box temperatures to the maintenance department in real time. To limit the impact of a pantograph failure on the cautionary system, an automatic monitoring facility which is currently being tested and will be developed shortly will detect any unusual localized wear on the contact strip.

Some of the measures taken by the design team in order to increase maintainability of the TGV train are given below:

- Maintainability of articulated trainsets: the articulated fixed formation structure of TGV trains is particularly well-suited to very high speeds, but originally in the design stages, there were some who feared that they would not be flexible enough in operation if vehicles had to be withdrawn for repairs. Five years of intensive operation and low-cost maintenance have clarified this debate. There have been very few occasions when a passenger vehicle had to be withdrawn and these have been detected prior to departure for a train service.

The situation would be no different with a non-articulated train; moreover, in this instance the many connections between cars and the complexity of inter-car gangways designed for high speeds would make it impossible to withdraw a car easily and prepare another, without creating major delays in the train service.

By contrast, maintainability of French TGV rolling stock and well-adapted terminal installations make it possible to place the full trainset back in service within a fairly short space of time by replacing a failed component instead of an entire car, even if it is an axle or a truck that has failed. Generally this principle is applied even to the end car that nevertheless is a separate vehicle in the articulated trainset.

At the Paris-Southeast workshops it takes about 1 h to change an axle and an idle truck in a trainset can be changed in no more than 1 h 30 min including all associated operations.

- Fast and accurate troubleshooting: the first generation of TGV trains was already equipped with testers together with diagrams of circuit logic which provided a good standard of troubleshooting. On Atlantic TGV trainsets, memory functions assigned to the various command and control circuit microprocessors store the values of operating parameters when a failure occurs. This facility provides a graph of failures, guarantees correct fault diagnosis and guards against recurrence of intermittent failures. The train-built computer and various test equipment form a comprehensive computer-aided fault diagnosis system, although it must be said that the number of failures should be minimal.

Preliminary organization measures:

- Documentation: it would be wrong to overlook an important aspect of maintainability; documentation, which is an essential pre-requisite for organization of maintenance. If supporting documents, drawings, technical descriptions, diagrams and functional manuals were not circulated prior to delivery of rolling stock, the equipment would not be maintainable and could not be placed in service without creating a risk. At the SNCF in general and especially for TGV trains, a large proportion of this information was made available well before lead times.

- Maintenance regulations: these documents are used to make a preliminary examination of initial maintenance duties and of expected maintenance intervals; maintenance specialists also use them to draw up their own maintenance manuals. Innovation rolling stock like the TGV is designed using components that have been tried and tested extensively for durability and the types of failures that are likely to occur are known. Hence, it is indeed relevant for teams of experienced engineers and technicians to foresee where failures might occur and draw up initial maintenance regulations. Of course, in the early period following commissioning, these regulations are more severe voluntarily, but they are very quickly adapted to the situation in practice and become cost-effective within a short space of time.
- Instruction and training: this dual-stage preparation of technical documentation specifying maintenance practices is used by future management teams already appointed to take charge of organizing supplementary training for staff selected for the home depot of trainsets. This is followed by a period in which maintenance duties are simulated and then by systematic training to perform these duties on sub-assemblies and on the first few trainsets delivered.

Although it is quite certain that system reliability and component durability play a key role in the quality of service and the economics of the TGV network, it is important to point out that all of these technical and organizational measures for trainset maintainability have been instrumental in ensuring simple and low-cost maintenance and for the reasonable level of unavailability.

Training in the new types of technology used was organized in advance and maintenance staff adapted smoothly to the problems encountered and to changes in qualifications.

- Maintainability of maintenance tools has contributed to ease of maintenance work, an element of comfort that is vital to the care and professionalism required for this work.

Maintainability and all of the measures connected with it form a new approach which calls for open and active relations among the partners involved and represents a challenge in terms of qualifications, organization and time-scales which the French railroad industry and the SNCF Rolling Stock Department have met successfully for TGV rolling stock.

Regarding the preventive maintenance tasks the following considerations were made:

- Automatic monitoring equipment is designed to meet the need to examine and inspect rolling stock regularly. Testing and fault detection equipment is designed to meet the need to re-establish redundancy promptly when temporary failures occur.
- Convenient location of items for ease of access.
- Items used for a particular technology were grouped together in functional units corresponding to the same technical specialty;

- Wear has been reduced some time by lubrication of moving mechanical items (gearing, roller bearings), and in some cases by replacing moving items with a solid-state technology. For example electro-mechanical switching and contact functions have been replaced advantageously by wear-free and maintenance-free static convertor power electronics.
- Ease of cleaning and possibilities for mechanized cleaning are also taken into account in the design of passenger stock for reasons of hygiene, comfort and aesthetics.

Regarding the corrective maintenance tasks some of considerations made at design stage are addressed below:

- Provision was made for testability, which practically means that the possibility of measuring the orders of magnitude of the physical parameters which are essential for fault detection, although it was not functionally necessary. Hence, many of the complex functions incorporated in the TGV include integrated test facilities or a remote fault detection system; these systems may function as fault analysis systems and include a facility to transmit data to repair centers.
- For maintenance tasks involving replacement of failed items, every provision is made for ensuring safety and swift replacement (snap-on mountings, polarized slots, lifting and handling gears and similar);
- The repair and renewal capacity of structures has been considered, *i.e.*, weldability, dismountability of items and parts vulnerable to impact, wear and ageing.
- Selection of materials and housings with objective to eliminate problems such as combustion, oxidation and ageing, which for decades represented the major part of repair and renewal work for railroad equipment; and
- It should be noted that the majority of systems failures have not been caused by the malfunction of some exotic device the design of which pressed the state-of-the-art. Rather, parts were not made correctly (bogus parts), and in other cases human failures such as failure to torque and secure a fastener properly or failure to install an explosive device properly. No detail is too minor to cause a problem. High inherent and achieved maintainability are, to a considerable degree, the system of painstaking attention to detail.

In this chapter the range of design responsibilities has been addressed, with a particular emphasis on maintainability requirements. The reasons why inherent maintainability must be high and why maintainability is needed as an independent check-and-balance function on design to assure that maintainability requirements and considerations get their prompt and proper share of design attention, have been presented. Further, the methods for designing a maintainable system were explored and the methods, procedures, and practices used in achieving and assuring maintainability in design were reviewed.

In summary, inherent maintainability is the primary responsibility of the design organization with maintainability service as an independent check and balance on the design function, principally to make sure that the design function has given its

maintainability responsibility detailed attention that is necessary. In addition, maintainability performs certain functions wherein its work is checked by design for the same reasons.

21.6.5 Maintainability Engineering Management – Lessons Learned

As a result of extended research through maintainability literature, design guidelines, and personal experience, the following selection of “lessons learned” has been created as a remainder and guide to the maintainability engineers of the future systems:

- Use of standard parts should be encouraged as far as possible, (seals, nuts and bolts, and all other high replacement rate items);
- Gaining access to items requiring maintenance, should not require removal of other items;
- Lubrication should be possible without disassembly;
- Relieving force in powerful springs should be possible before they can be removed;
- Fitting pipes to items should be in one axis so that the item can be removed in one direction;
- Items which come into contact with tools should not be painted;
- Adequate wall thickness should be provided if a hole in a body is used for lock-wire in order to prevent breakout after repeated use;
- Labels and decals should be hard wearing, not fading, positioned to avoid damage, difficult to peel off, easy to renew, and should follow contours of item without lifting;
- Items not visible after assembly should be retained to prevent their dislodging during the assembly of other items;
- Positive indication of locking together of items should be provided;
- Location of a bolt/fastener should be shown by an indicator arrow if its location is not easily visible;
- Pointed bolts should be used when alignment may be difficult;
- Small thread sizes should be avoided as they are prone to damage;
- Good access should be provided to any item which require torque loading;
- Use of special tools should be avoided;
- Provision of visible indication of correct installation of critical items should be provided;
- Cables needing inspection and repair should be easily accessible;
- Cables should be secured with reusable clips;
- Cables should be kept to a minimum practical length in order to minimise the risk of damage due to excess slack;
- Plugs and sockets should be identified by shape, colour coding, or similar means;
- Possible galvanic reaction between dissimilar metals should be considered (stainless steel and aluminium mating should be avoided);
- Attachment of test equipment should be provided for in-place testing;

- Conformal coating of the PCB should be considered, particularly in cases when it is not environmentally sealed;
- Environmental and EMC seals between mating metal surfaces should be considered especially for safety critical items;
- Use of existing test equipment should be encouraged where practical;
- items should be tested in their completed form with no need for subsequent further assembly;
- Adjustments on the item on the test rig should be possible, obviating the need for removal, estimated adjustments, and subsequent retests;
- Sufficient clearance to remove and refit a seal without causing damage to it or the item to which it is fitted should be provided; and
- Automatic renewal of old seals should be ensured if new seals are needed with a new part.

Regarding assembly of units:

- Components not visible after assembly should be retained to prevent their becoming dislodged during the assembly of other components;
- Bias indicators in one direction to give an unambiguous indication of status *e.g.*, plungers on gas sears should be sprung loaded so that they cannot be left in the cocked position unless the component is cocked;
- A positive indication of locking together of components should be provided *e.g.*, seat locked to gun, trombone tubes to manifolds “flush for correct” should be used throughout the design;
- Where the location of a bolt or fastener is not easily visible an indicator arrow should show its location;
- Use a full hexagon instead of two flats to allow maximum access and engagement;
- Avoid the need for special peg spanners *e.g.*, use hexagons; and
- If peg spanners are needed, use existing ones where possible, with the maximum practical hole centers and peg diameters.

Regarding fasteners:

- Where flats are used on components for removal use one standard size throughout;
- Where alignment may be difficult use pointed bolts;
- Avoid left-handed threads;
- Use correct fasteners ensuring correct shank length on structural items;
- If a bolt needs to be cropped consider using a nylon bubble nut;
- Use Quick Release (QR) pins where possible or pins needing a simple tool to release if there is a risk of QR pins becoming dislodged;
- Avoid the use of circlips. If circlips are necessary use ones which all require the same pliers size to remove;
- Avoid the use of split pins especially in poor access areas. Do not recess the split pin even by only $\frac{1}{4}$ ”;

- All nuts and bolts should not be recessed so that they can be started by hand rather than balanced in a socket or spanner;
- There should be sufficient access around nuts and bolts for standard tools; they should not require ground down spanners;
- Maximum use should be made of one type of nut and bolt for each item;
- In general use hexagon headed bolts – in poor access area socket headed bolts should be considered;
- Avoid Hi-Torq and Torq-Set bolts as they require special screwdrivers that are prone to damage;
- With high removal rate items, use captive bolts if possible or as a last resort, helical inserts;
- Secure pipes by the mounting of the unit rather than by separate nuts and bolts, or use plug-in connections secured with a bolt in preference to flare nut fittings;
- Avoid small thread sizes as they are prone to damage or stripping.
- It should be possible to fit bolts all the way through and then fit the nut, having to put the bolt part way through, fit the nut, and tighten the bolt is unacceptable;
- Any items that may require torque loading should have especially good access with a socket rather than a spanner;
- The need for special tools and GSE should be avoided; if FSE is required the component would be designed to use existing items; if new items have to be developed then the new item should be multi-use, *e.g.*, all timing mechanisms should accept the same cocking tool;
- Use spring clips rather than p-clips for items that will be removed relatively frequently;
- Environmentally seal units were practical and where this would result in better reliability; and
- Provide visible indication of correct installation of critical components.

Regarding electrical cables – installation:

- Cables should be routed so as to be readily accessible for inspection and repair;
- Secure cable with reusable clips rather than cable ties;
- Do not route over items that are removed for routine maintenance; and
- Avoid routing cables adjacent to or across ballistic gas pipes, if absolutely necessary ensure suitable protection is used.

Regarding electrical cables – design:

- Keep cables to a minimum practical length to minimise risk of damaged due to excess slack and so that they cannot be wrongly fitted when all the cables are connected;

- Quick disconnect plugs that require no more than one turn to fully lock them in position should be used. Use MIL-C-38999 series III connects or equivalent wherever possible;
- Plugs should use aligning pins that extend beyond the electrical pins and have their location marked on the plug body;
- Shape, colour coding and similar means should identify plugs and sockets;
- Allow sufficient spacing between adjacent connector to allow room for locking and unlocking;
- Plugs should have visual indication of a correctly locked connection.
- Wherever a connection is required, the connector with sockets should be used for the live side and pins for the non-live side;
- In allocating each wire's pin or socket position in a connector, due consideration should be given to the function of that wire and its voltage level and frequency;
- Segregate pints with high relative voltage and group together related pins such as signals and returns;
- If there are bonding and earthing requirements then connectors with "EMC fingers" should be used; and
- Consideration should be given to the possible galvanic reaction between dissimilar metals, *e.g.*, do not mix stainless steel and aluminium mating connectors.

Regarding electronic modules:

- Modules should be able to be tested in place with sufficient access allowed for attachment of test equipment;
- Printed circuit boards (PCBs) should be manufactured and released to the requirements of MIL-P-55110 type 3. PCB artwork, wiring details and mechanical outline constraints should be in accordance with MIL-STD-275E;
- Fastening points on the PCB should not be directly above embedded tracks;
- Consideration should be given to conformal coating of the PCB particularly if the case is not environmentally sealed;
- The incorporation of environmental and EMC seals between mating metal surfaces should be considered especially for safety critical applications; and
- To improve the reliability of service equipment a Burn-In program should be put in place. This generally consists of a combination of random vibration and temperature cycling.

Regarding testing:

- Use existing test equipment where practical, fitting adaptors where necessary;
- Components should be tested in their completed form with no need for subsequent further assembly;
- If an assembly needs to be reset after testing, it must not be necessary to dismantle any part of it to achieve this; and

- It should be possible to make all necessary adjustments with the component in place on the test rig, obviating the need for removal, estimated adjustments, and subsequent retests.

Regarding seals:

- Ensure that there is sufficient clearance to remove the refit a seal without causing damaged to it or the component to which it is fitted;
- If new seals are needed with a new part, have the seal fitted to the part so that renewal is automatic; and
- Avoid the use of internal 'O' seals unless of a reasonably large size and with good access.

Regarding lubricants:

- Use only one type of oil and grease throughout;
- Avoid differing types of 'O' seal material that require different types of grease;
- Consider dry film lubricant in exposed areas; and
- Use self-lubricating or sealed for life bearings where practical.

21.7 Concluding Remarks

The main purpose of existence of any man-made item/system is the provision of utility by performing a required function with expected performance and attributes. Hence, once the functionality is provided, the main concern of the user is to achieve the highest possible functionality and safety at least possible investment in resources.

Performance of any maintenance tasks is related to associate costs, both in terms of the cost of maintenance resources and the cost of the consequences of not having the system available for operation. Therefore, maintenance departments are one of the major cost centers, costing industry billions of pounds each year, and as such they have become a critical factor in the profitability equation of many companies'. Thus, as maintenance actions are becoming increasingly costly, the maintainability engineering is gaining recognition day by day.

It is clear, from the brief analysis of the role and the importance of maintainability given above, that it represents one of the main drivers in achieving user's goals regarding functionality, reliability, cost of ownership, reputation, and similar objectives. For example, in the Journal Aviation Week & Space Technology, of 22 January 1996, it was reported that by year 2000 the USA Air Force will begin looking at upgrades of a heavy air-lifter aircraft C-5A. The comment was that, although the structure of aircraft is considered to be good, "the reliability and maintainability leaves a lot to be desired". It is inevitable that in the future considerations and comments like this will significantly increase and that the impact of these considerations on the final selection of the systems will be far greater.

Thus, the analysis of concepts, tools, techniques, and models, available to the maintainability specialists for the prediction, assessment and improvement of their decisions related to the ease, accuracy, safety, and economy, of performing tasks related to maintaining systems in a functionable state during their utilization, which directly influence the length of time which system will spend in SoFa, are the main concerns of this chapter.

To round up this introductory part, the final example used is related to working practices applied during the creation of a new Boeing 777. It is based on the private communication from the author with Mr. Eugene Melnick, Maintainability Engineer from Boeing Corporation, Seattle.

The 777 Airplane has been designed for a useful life of 20 years. Boeing recommends and authorities of the FAA and JAA decide what maintenance is required to keep the airplane airworthy while in service. This involves defining what minimum scheduled and unscheduled maintenance must be performed in order to continue flying. Scheduled maintenance is performed at certain intervals that are tied to number of flight hours, number of cycles (such as turn-on/off, take-offs and landings), *etc.* It consists primarily of inspections followed by maintenance, corrosion prevention, *etc.* Unscheduled maintenance is performed after a failure occurs. Depending upon the criticality of the failure, maintenance is accomplished either before the airplane is returned to revenue service or within a specified interval.

When total cost is considered over the life cycle, it is evident that the operating and support costs of the airplane will eventually exceed the initial acquisition cost. In order for Boeing to make the airplane attractive to the airlines, the engineers must include maintenance cost savings in the design. Reliability and maintainability issues addressed this. Increased reliability means fewer failures to fix. Increased maintainability means shorter maintenance times.

The figure of merit chosen to measure reduction of the follow-on costs was schedule reliability. In other words, how often will the airplane, or fleet of airplanes, meet the scheduled take-off time? The target for initial delivery is 97.8% with improvement to 98.8% after fleet maturity. In order for the airplane to meet such a high standard, it must be inherently reliable. Double and triple redundancy is used in critical areas, allowing deferral of maintenance to an overnight time while the back-up system or systems keep the plane flying until that time.

Maintenance must be able to be completed during the scheduled downtimes, whether it is during a 45-min turnaround between flights or during an overnight. This implies that there are good means of identification and isolation of failures, as well as good access to the equipment. Innovative computer aided human models were used to prove good maintenance access without the use of expensive mock-ups. Fault identification and isolation is enhanced with the use of extensive built in testing with fault messages displayed on the computer screens available to the mechanics. Great care was used to ensure that maintenance messages are prioritized, understandable, do not give extraneous information, and are accurate. Accompanying fault isolation and maintenance manuals complement this information.

Reliability requirements were passed along to equipment manufacturers by specifying the mean time between failures (MTBF) and target mean time between

unscheduled removals (MTBUR). The latter was estimated to be between 0.8 and 0.9 of MTBF, but could be verified only by service experience. It was recognized that unscheduled removals also counted the times that equipment is wrongfully removed because of the haste that a gate mechanic expends in trying to clear a fault during a 45-min turnaround. The tendency is to replace the first suspected unit or groups of units in order to eliminate the obvious faults from the process. Thus the maintenance messages must give the right information that avoids removing good items. Specifying both MTBF and MTBUR means both inherent reliability and field reliability could be controlled.

For fault tolerant systems or items, the reliability index was mean time between maintenance alerts (MTBMA). Maintenance alerts are the maintenance messages that are documented on equipment with internal failures that did not immediately affect function.

Boeing also documented 'lessons learned' data to record service history and feedback from other airplanes in order to avoid the same mistakes in the design of the new airplane. The airline representatives stayed in touch by attending design reviews and other meetings of concurrent engineering teams. From time to time, their field mechanics visited Boeing to provide their inputs. The result was a working-team relationship that benefited both sides and will result in increased reliability and maintainability.

It is necessary to stress that the distribution approach to maintainability analysis does not required any additional testing time that practically means that all additional information could be obtained at no extra testing cost.

In the days ahead, when the investment in the resources needed for the operation and maintenance of modern and complex equipment will be restricted, even reduced, the required level of maintainability/availability will be in greater demand. Consequently, the effectiveness of the approach selected for the analysis of the maintainability data will have an important role to play.

The chapter also demonstrates the methodology for the fast and accurate prediction of maintainability measures of the future maintenance tasks and the resources needed for their completion. The proposed method is based on the maintenance block diagram and maintainability measures related to the comprising activities.

The method proposed is applicable to maintenance tasks whose consisting activities are performed simultaneously, sequentially, and combined. Thus, it is a generic model for the fast prediction of maintainability measures, which in return represent one of the most important preconditions for the successful completion of logistic support analysis. It is necessary to stress that the method presented could be successfully used at the very early stage of design when most of the information available is based on previous experience, as well as in the detailed and test stage when the relevant data are obtained from the adopted design solution.

The chapter stresses the need for maintainability engineering function that may also include the coordination of the design test programs, direction of independent maintainability test programs and actual conduct of all testing, identification and establishment of control systems for portions of the design which have special limitations, preparation of maintainability specifications applicable to suppliers,

and the imposition of maintainability requirements on suppliers through review and approval of procurement documents.

References

- Blanchard BS and Lowery EE (1969) Maintainability Principles and Practices, McGraw-Hill Book Co., New York.
- Blanchard BS, Verma D, Peterson EL (1995) Maintainability, John Wiley & Sons, New York.
- Cox DR (1962) Renewal Theory, Methuen & Co. Ltd, London, U.K.
- Dubi A (2003) System Engineering Science, Analytical Principles and Monte Carlo Methods, pp. 168, MIRCE Science, UK.
- Einstein A (1991) The World I See It, translated by Alan Harris, Citadel Press Book, New York, U.S.A.
- Gleick J (1993) Genius. Panthenon Books, New York.
- Hunt S (1993) Time Dependent Maintainability Analysis, Proc. International Symposium, Advances in Logistics, edited by J. Knezevic, Research Centre MIRCE, Exeter, U.K.
- Knezevic J (1985) Methodology 'EXETER' for Selecting an Optimal Part-Replacement Policy. Maintenance Management Int., 5: 209–218
- Knezevic J (1993) Reliability, Maintainability and Supportability Engineering – A probabilistic Approach, pp. 292, plus software PROBCAR, McGraw Hill, London, U.K.
- Knezevic J (1994) Maintainability Prediction at the Design Stage, Journal of Communications in Reliability, Maintainability and Supportability, SAE, 1(1): 24–29.
- Knezevic J (1996) System Maintainability – Analysis, Engineering and Management, pp 460, Chapman Hall, London, UK.
- MIL-STD-470B (1989) Military Standard, Maintainability Program for Systems and Equipment, Department of Defence, Washington, D.C., USA.
- MIL-STD-721B (1966) Military Standard, Definitions of Effectiveness Terms for Reliability, Maintainability, Human Factors and Safety, Department of Defence, Washington, D.C., USA.

Part VII

Maintenance Safety, Environment and Human Error

Safety and Maintenance

Liliane Pintelon and Peter N. Muchiri

22.1 Setting the Scene

The desire to be safe and secure has always been an intimate part of human nature since the dawn of human history. The demand for safety and security is pursued at every location in one's entire environment. This ranges from homes, in transit, at all premises, and indeed; in the workplace. The need for a safe working environment was first brought to light during the first decade of industrial revolution (Roland and Moriarty, 1983). Based on the knowledge acquired in the past decades, companies and labour organizations have pursued ways and means of enhancing occupational safety. Since 1950, the International Labour Organization (ILO) and the World Health Organization (WHO) have had a common definition of occupational health and safety. This definition was adopted by the Joint ILO/WHO Committee on Occupational Health at its First Session (1950) and revised at its 12th Session (1995):

“Occupational health should aim at: the promotion and maintenance of the highest degree of physical, mental and social well-being of workers in all occupations; the prevention amongst workers of departures from health caused by their working conditions; the protection of workers in their employment from risks resulting from factors adverse to health; the placing and maintenance of the worker in an occupational environment adapted to his physiological and psychological capabilities; and, to summarize, the adaptation of work to man and of each man to his job” (Source: www.wikipedia.com).

However, the road to enhancing occupational safety has not been as smooth as some statistics tells us. According to Hammer, 8% of the workers in the US suffer some kind of accident at work each year, though few involve disabilities or death (Hammer, 1976). Though these statistics were recorded 30 years ago and may appear to be old, the recent statistics indicates persistence of occupational safety problems. Work-related death still occurs regularly and in 1996, the National Safety Council of US estimated 3.9 million disabling injuries, 4,800 deaths and the

total cost of work-related deaths and injuries was estimated to be \$121 billion annually (NSC, 1997). This prompts an important question that begs for answers: why do accidents happen in the workplace, and what can be done to prevent them? These questions have troubled plant engineers and managers for decades, and have led to a substantial increase in safety knowledge and accident prevention and investigation activities in all industries. Government agencies, which enforce statutes and regulations, have also added to safety awareness and action based on regulation, inspection, and penalties.

Despite increase in safety knowledge, there has been an increase in production system automation so that both operational and safety-related equipment is more complex to understand and properly maintain. These improperly maintained or unmaintained pieces of equipments pose a major safety hazards to the plant. Moreover, the autonomous maintenance movement involves operators in certain maintenance tasks, exposing them to more potential hazards. No doubt, the impact of maintenance on plant safety has never been so significant. Maintenance in many industries is connected with a significant proportion of the serious accidents occurring in the industry. Studies by the British Health and Safety Executive (HSE, 1987) of the deaths in the chemical industry showed that some 30% were linked to maintenance activities, taking place either during maintenance activities or as a result of faulty maintenance. A study of the chemical accidents stored in the database FACTS (Koehorst, 1989) found 38.5% of the accidents where dangerous materials were released from on-site plant had taken place during maintenance. Another study by Hurst on 900 accidents involving pipe-work failure in Chemical plants found out that 38.7% have their origins in the maintenance phase of plant operations (Hurst *et al.* 1991).

Maintenance function forms an integral part of manufacturing and its price tag can possibly indicate its significance to manufacturing plants. A study conducted in 1999 indicates that United States spends \$300 billion on plant maintenance and operations (Latino, 1999). As billions are being spent each year on maintenance to keep engineering systems and items in operational state, the problem of safety in maintenance has become an important issue (Dhillon, 2002). Some examples in practice prove the importance of this topic. Report from the National Safety Council of US shows that in the mining industry, 13.61% of all accidents occurred during maintenance in 1994 and, since 1990, the occurrence of such accidents has been increasing each year (NSC, 1999). A study of electronic equipment revealed that approximately 30% of failures were caused by human error with faulty maintenance contributing 8% of the failures (US Army, 1972). Another study carried out between 1982 and 1991 on safety issues with respect to onboard fatality of worldwide jet fleet revealed that maintenance and inspection was the second was the second most important safety issue with a total of 1481 on board fatalities (Russell, 1994).

The questions that arise from these statistics are; what is the impact of maintenance (and the corresponding policies) on plant safety, how do maintenance jobs interact with safety (or create safety hazards)? what interventions (technical, managerial, legal) can be employed to improve plant safety? The study on maintenance and safety interactions in industries, therefore, is indispensable. It is our objective first to establish a link between safety and maintenance. This will be done by looking at the safety issues in maintenance work and maintenance for

safety of production equipments. The second objective is to study the effect of various maintenance policies and concepts on plant safety. The third objective is to study how safety performance can be measured or quantified. This will be coupled by cost and benefit analysis of safety improvement efforts on the plant. Finally, accident prevention will be discussed in reflection to the safety legislation put in place by governments and some safety organizations. Let us look at some definitions and terminologies used in this study.

22.2 Definitions

22.2.1 Maintenance

The British Standards Institution (BSI, 1984) defines maintenance as:

“A combination of all technical and associated administrative activities required to keep an equipment, installations and other physical assets in the desired operating condition or restore them to this condition”.

Though this is what maintenance indeed is, it would be confirmed by any practitioner, its role could well be defined by the four objectives it seeks to accomplish. These are (1) ensuring system function (availability, efficiency and product quality), (2) ensuring the system or the plant life, (3) ensuring human well-being and, finally, (4) ensuring safety (Dekker, 1996).

For production equipment, ensuring the system function is the prime objective of maintenance function. Here, maintenance has to provide the right reliability, availability, efficiency and capability to produce at the right quality for the production system, in accordance with the need for these characteristics. Ensuring system life refers to keeping systems in proper working condition, reducing chance of condition deterioration, and thereby increasing the system life. Maintenance for ensuring human well-being or equipment shine has no direct economical or technical necessity but primarily a psychological one of ensuring the equipment or asset looks good. A good example is painting for aesthetic reasons.

The last but very important objective of maintenance is to ensure safety of production equipments and all assets in general. As explained by Hale *et al.* (1998) the primary purpose of maintenance is to prevent significant deterioration or deviation in plant functioning, which can threaten not only production but also safety and to return a plant to full functioning after breakdown or disturbance. While maintenance function seeks to ensure safety of the plant, many maintenance tasks expose maintenance staff to potential safety hazards. No doubt the maintenance function has a significant impact on the plant safety.

22.2.2 Safety

Based on the dictionary, safety is as the condition of being free from undergoing or causing hurt, injury, or loss. It is freedom from any potential harm. The standard definition is:

“Safety is defined as the condition of being free from or protected against failure, damage, error, accidents, or harm or any other event, which could be considered undesirable (Wikipedia- <http://en.wikipedia.org/wiki/Safety>). Safety in a system is defined as a quality of a system that allows the system to function under predetermined conditions with acceptable minimum of accidental loss” (Roland and Moriarty 1983).

22.2.3 Hazard

Safety is generally interpreted as implying a real and significant impact on risk of death, injury *or* damage to property. Lack of safety occurs due to existence of hazards in the workplace.

“A hazard is defined as any existing or potential condition in the workplace which, by itself or interacting with other variables, can result in the unwanted effects of death, injuries, property damage, or other losses” (Laing, 1992).

22.2.4 Stimuli

The presence of a hazard by itself cannot directly lead to an accident. A trigger is needed convert a hazard to an accident. This trigger is known as stimuli:

“Stimuli is defined as a set of events or conditions that transforms a hazard from its potential state to one that causes harm to the system, related property or personnel”(Roland, 1983).

22.2.5 Accident

Accident is the outcome of a hazard that is triggered by a stimuli. An accident happens when there is loss of plant system or part of the system, injury to or fatality of the operators or personnel in near proximity, and property damage of related equipment or hardware. Therefore:

“An accident is defined a dynamic mechanism that begins that begins with the activation of a hazard and flows through a system as a series of events, in a logical sequent, to produce a loss.”

Risk is associated with likelihood or possibility of harm or the expected value of loss. Risk is related to the probability that frequency, intensity, and duration of

the stimulus will be sufficient to transfer the hazard from potential state to a loss. Having defined the maintenance, safety and workplace hazards, the next thing that arises is to see how these issues relate to or interact with each other in the workplace.

22.3 The Maintenance Link to Safety

22.3.1 The Role of Maintenance

Maintenance has a major relevance to the business performance of industry. Whenever a machine stops due to a breakdown, or for essential routine maintenance, it incurs a cost. The cost may simply be the costs of labour and the cost of any materials, or it may be much higher if the stoppage disrupts production.

In many instances, the production pressures to meet the production targets are very high in the manufacturing environment. Maintenance is pressurized to ensure plant's availability and to support the desired output. In many manufacturing plants the question is *production or maintenance*? Faced with the choice of running full tilt or halting for scheduled upkeep, plant managers typically have the upper hand over their maintenance colleagues and opt for production. The latter can be a costly choice and may be detrimental to the production process and to the plant's safety.

The importance of maintenance to manufacturing can be termed as paradoxical. This is because, when breakdown happens, it is often easy to show that lack of maintenance was responsible. Nevertheless, when there is no breakdown, it is not easy to demonstrate that maintenance had prevented them. This is due to the traditional attitude of production management towards maintenance as a non-productive support function and as a necessary evil (Pintelon *et al.* 1997). This attitude towards maintenance may have serious consequences for plant safety.

Maintenance actions, objectives and strategies are influenced by the company policy, sales and production policies, and other conflicting demands and constraints in the company. Maintenance resources are utilised so that the plant achieves its design life, so that safety standards are achieved, so that production volume required by production policy is met and so that energy use and raw material consumption are optimised among other factors. All these factors influence the maintenance objectives as shown in Figure 22.1.

In spite of its contradictory relationship to manufacturing, maintenance is a very important function in a plant's operating life. As soon as the plant is commissioned, deterioration begins to take place in the components. In addition to the normal wear and deterioration, other failures may also occur when the equipment is pushed beyond its design capacity. Degradation in equipment condition results not only in reduced equipment capability but also in undesirable safety condition. Without regular inspection and maintenance, plant and machinery soon or later lapse into a dangerous state due to wear, tear, fatigue and sometimes corrosion. Regular inspection is therefore needed to determine in detail how far such deterioration has proceeded. This is done many times against production pressures that demand to meet certain production targets.

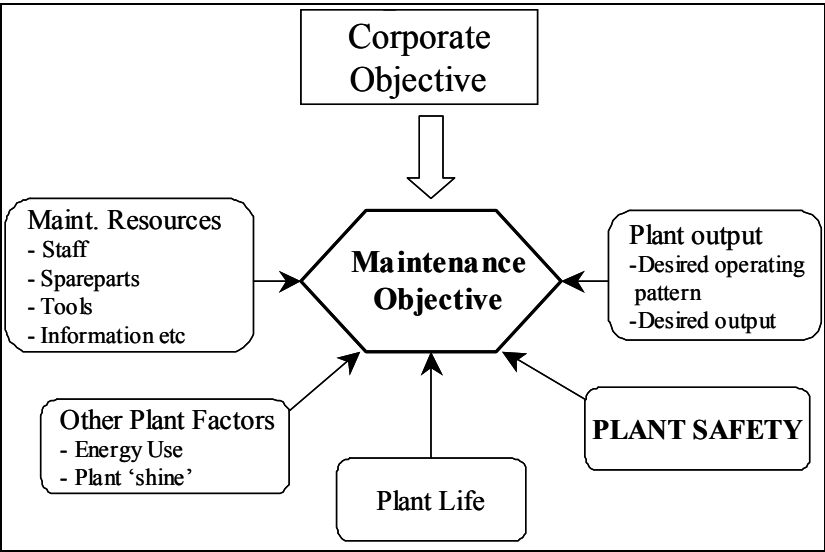


Figure 22.1. Factors influencing the maintenance objectives

Most maintenance activities require the plant or equipment be shut down and specially prepared. Consequently, a minor job (often referred to as a repair) or a major job (often referred to as an overhaul) is carried out. A limited and clearly defined number of maintenance jobs are done while the plant is still running. A good example of these are the traditional lubrication as well as advanced condition-based monitoring. Maintenance may be planned or unplanned. Unplanned or breakdown maintenance means operating the plant until something breaks down. The break down may be classified as emergency or corrective depending on its urgency although the work done may be the same in both cases. Planned maintenance involves both preventive maintenance and corrective maintenance. Preventive maintenance is based on servicing and overhauling key plant items before they breakdown or their performance deteriorates. This is done at pre-selected intervals dependent on the equipment usage and is therefore referred to as use-based or scheduled maintenance. Condition-based or predictive maintenance is also used to monitor the condition of equipment to proactively correct undesirable condition. Maintenance function can be summarised as in Figure 22.22.

The primary purpose of both preventive and corrective maintenance is to prevent significant deterioration of or deviation in plant functioning, which can threaten not only production but also safety of the plant and to return a plant to full functioning after a breakdown or disturbance. If maintenance is not carried out soon enough, is incorrectly carried out, or communications between maintenance and operation staff are not effective, the plant may fail dangerously during start up or during normal operation phase (Hale *et al.* 1998). An example of maintenance related accident is Piper Alpha company in 1988 with 165 fatalities among many other examples (Dept of Energy, 1990). In addition, the autonomous maintenance movement involves operators in certain maintenance tasks. Many of these tasks

have considerable risks and therefore expose the maintenance staff to more potential hazards. Some examples of accidents happening during maintenance work is Phillips in Pasadena in 1989 with 23 fatalities and Arco in Texas in 1990 with 17 fatalities (Craft, 1991). Though it is apparent that maintenance has a considerable impact on plant safety, little has been written about maintenance interaction with safety. The first attempt to investigate the impact of maintenance function on plant safety was done by Ray *et al.* (2000). Their study showed an inverse relationship of moderate strength between injury frequency index and maintenance audit score. The finding of the study supports the hypothesis that better maintenance, as presented by better audit score, is associated with lower injury frequency.

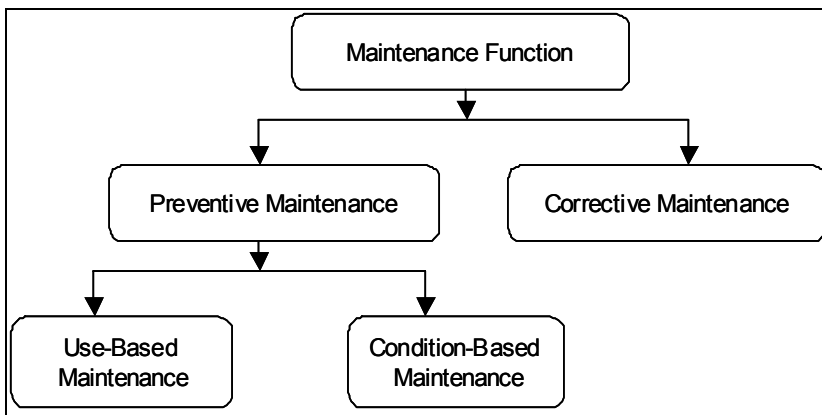


Figure 22.2. Maintenance function layout in plants

We can classify the hazards connected with maintenance job into three categories:

1. Hazards occurring during maintenance;
2. Hazards caused by faulty maintenance; and
3. Hazards caused by lack of maintenance.

Hazards of type 1 occur while maintenance is taking place. Accidents of this type may occur for several reasons, among them being maintenance working under production pressure, high workload, failure to follow the required procedures, complex technology, shortage of skills, lack of expertise, a cut in maintenance budget, insufficient maintenance facilities, lack of spares or lack of support from top management. These factors raise a concern on “*safety during maintenance*” that will be covered in Section 22.2 below.

Hazards of type 2 and 3 occur when maintenance is not carried out appropriately or when maintenance intervention is not done at the appropriate time. This raises an issue of “*maintenance for safety*” that will be covered in Section 22.3.2.

22.3.2 Safety During Maintenance

Maintenance work may significantly increase the likelihood of work injuries across many industries. Because of the nature of maintenance work, craftspeople are usually over-represented in the group of injured workers – regardless of industry or level of aggregation of accident statistics (Batson *et al.* 1999). In some organizations, the maintenance people have the highest injury rates and furthermore have the highest exposure to hazardous chemicals (Levitt, 1997). A recent study carried out in France (Pichot, 2006) followed 1,250 maintenance workers for 5 years (1995–2000). The study revealed that maintenance workers were 8–10 times more vulnerable to occupational diseases than other workers. The accident rate in maintenance was slightly smaller than the national average. However, for some maintenance specialities, this accident rate was much higher than the average. The accident severity of maintenance accidents measured in days away from work is 29% higher than the national average.

When the Occupational Safety and Health Administration (OSHA) promulgated its Lockout/Tag out Standard in 1989, the agency estimated that 122 fatalities, 28,400 lost workday injuries and 31,900 non-lost workday injuries resulted each year from accidents involving the maintenance, repair, or servicing of equipment. Almost 75% of these accidents occurred in manufacturing facilities. Most (88%) of the injuries were caused by moving machine parts, with agitators and mixers, rolls and rollers, conveyors and augers, saws and cutters, and hoists accounting for 63% of the fatalities (OSHA, 1989). To support these estimates, OSHA cited a Bureau of Labour Statistics survey of 883 workers injured while cleaning, un-jamming or performing other non-operating tasks on machines, equipments or electrical systems. According to this study, 74% of the accidents occurred in manufacturing industries; moving parts were cause of 88% of injuries. The occupational distribution of injured workers was operators, 45%, craft workers, 24% and mechanics and repairers, 10%. OSHA also reviewed 83 fatality investigations conducted between 1974 and 1980; 25% of these deaths were attributed to lack of adherence to safe work procedures and 60% were caused by failure to properly de-energize machines and equipments before performing maintenance. Agitators, mixers, rolls and rollers, conveyors, augers, saws and hoist were involved in 63% of fatalities (OSHA, 1989).

Batson *et al.* (1999) also quote some statistics from empirical evidence from a local automotive component plant that showed that maintenance personnel had 13.9% of the injuries in a past year. Two of the four operational departments had 27.1% and 18.3% of the injuries, respectively and operations accounted for 65% of the total. However, the accident rate for the maintenance department was higher than operations because there are significantly fewer maintenance workers than operators in this plant. They hypothesize that there is some validity to the claim that maintenance work can be the most dangerous work in a plant and that maintenance workers are involved in accidents at a rate that exceeds any other plant job classification.

Example 22.1: A gas explosion at a steel factory in Belgium drew attention to the hazards of maintenance work, especially maintenance work carried out by contractors. Two workers from a contractor firm died while replacing a valve in a production line. The part of the installation they needed to work on was supposed to be empty, but was not. When they started working a gas explosion occurred. It killed the 2 workers, 13 others were very badly injured and 13 more were lightly injured. A court investigation convicted two supervisors of the factory, because they gave the clearance to work on the line without properly checking if everything was okay.

These past statistics and experiences indicate that there are significant proportions of accidents that occur during maintenance and the nature of maintenance work exposes those who perform it to greater hazards. The question that arises from these statistics is what are the reasons for safety problems in maintenance and what factors are responsible for the dubious safety reputation in maintenance work. Dhillon (2002) outlines some important reasons for safety related problems in maintenance as follows:

- Inadequate equipment design;
- Poor work environment;
- Inadequate safety standards and tools;
- Poor management;
- Inadequate training to maintenance personnel;
- Poorly written maintenance procedures and instructions;
- Inadequate work tools; and
- Insufficient time to perform required maintenance task.

Stoneham (1998) also outlines the factors that make maintenance have a dubious safety reputation as follows:

- Performance of maintenance tasks in remote locations, at odd hours and in small numbers;
- Difficulty in keeping regular communication with workers involved in maintenance tasks;
- Sudden requirement for maintenance work, thus allowing a limited time for preparation;
- Frequent occurrence of numerous maintenance tasks and thus fewer opportunities for discerning safety-associated problems and for introducing remedial measures;
- Disassembling previously operating items, thus working under the risk of releasing stored energy;
- Need to carry bulky and heavy items from a warehouse or store to the maintenance site, sometimes using lifting and transport equipments way beyond the boundaries of a strict maintenance regime;
- Performance of maintenance work inside or underneath items such as large rotating machines, pressure vessels and air ducts;
- Time to time maintenance work may require carrying out tasks such as manhandling cumbersome heavy items in poorly lit areas and confined spaces or disassembling corroded parts; and

- Maintenance tasks performed in unfamiliar territories or surroundings imply that hazards such as broken light fittings, rusted handrails and missing gratings may go unnoticed.

22.3.3 Maintenance for Safety

The interaction between maintenance and safety goes beyond the simple occurrence of accidents during the conduct of maintenance work, however dramatic these may be, as explained in Section 22.2 above. The primary purpose of maintenance is to prevent significant deterioration of plant condition, which threatens not only production but also plant safety. If maintenance is not carried out in good time or is incorrectly carried out, the system can fail dangerously causing deaths, injuries and extensive destruction of property. Some facts, figures and examples can prove this:

- In 1979 in a DC-10 aircraft accident in Chicago, 272 persons lost their lives because of incorrect procedures followed by maintenance personnel (Christensen and Howard 1981).
- An incident involving the blow out preventor (assembly of valves) at the Ekofish Oil field in the North Sea was due to upside-down installation of the device and its estimated cost was around \$50 million (Christensen, 1981).
- In 1990, a newly replaced windscreen of a British BAC 1-11 jet blew out as the aircraft was climbing to its cruising altitude because of incorrect installation of the windscreen by a maintenance worker (Transport Ministry, 1992).
- In 1991, an explosion killed four people in an oil refining company in Louisiana. The explosion occurred as three gasoline synthesizing units were being put into operation after some maintenance activities (Goetsch, 1996).
- In 1983, three engines of a L-1011 Lockheed jet failed in flight after oil leaked from the engines because during routine maintenance, the maintenance workers overlooked the fitting of O-ring seals onto master chip detectors (Safety Board, 1984).
- In 1990, ten fatalities occurred on U.S.S. IWO Jima (LPH2) naval ship due to a steam leak in the fire room. An investigation into the accident revealed that maintenance workers just repaired a valve and replaced bonnet fasteners with mismatched and wrong material (US.Navy, 1992).
- In 1985, 520 people lost their lives in a Japan Airline Boeing 747 jet accident due to an improper repair (Gero, 1993).

There are many reasons quoted in the literature that cause these maintenance related accidents to happen. As noted by Batson *et al.* (1999) maintenance related failures are not intentional. Often maintenance management does not have safety standards in place, or has not trained their workers in safe maintenance practices. Moreover, the maintenance workforce may be overburdened with corrective maintenance or even paperwork related to work-orders, and their preventive maintenance schedule of activities is ignored or delayed. Tools needed for certain

adjustments may not be available; parts (e.g., replacement hoist or conveyor belts) at the time of scheduled replacement. Worse still, there may not even be a preventive maintenance program in place in some cases causing serious deterioration of some critical parts of the plant. This may be due to failure to inspect, detect and replace worn out parts, failure to lubricate equipments on scheduled basis or failure to tag and/or lockout unsafe equipments among others. The situation is complicated in many cases if tight production schedules are given higher priority than maintenance. In this situation, maintenance would only be carried out when time is available and thus the condition of the machinery is compromised. This may have serious consequences on the plant safety.

A study carried out to evaluate safety in the management of maintenance activities in the chemical process industry in the Netherlands (Hale *et al.* 1998) made far reaching conclusions on the causes of accidents and suggested areas in maintenance- safety management where attention is needed. From the study:

- It was estimated that around 40% of serious accidents in industries are related to maintenance, 80% of those occurring during maintenance phase and 20% in normal operations because of deficiencies in maintenance management. This confirms why there is a greater accident risks (often more than five times higher) for contractor's personnel compared to the own personnel.
- It was identified that there is a great weakness in the translation of general safety policy objectives into maintenance concepts, designs, planning, procedures and resource management to achieve improved safety. This translation process is the responsibility of senior management to support the function of the middle management and maintenance workers.
- It was noted that there is failure to incorporate safety into existing maintenance management systems as both exist in industries as independent functions.
- There is lack of a strong maintenance engineering function whose task is to coordinate the information flow between life cycle phases to ensure feedback in the previous plant experience; Thus, identify plant items responsible for accidents, incidents, breakdowns and problems in preventive maintenance; analyse root cause of these events; develop and implement improvements to prevent them by plant modifications, adjustment of maintenance concepts, operator training, *etc.* This would be important to improve the inherent safety and reliability of the plant and develop the most cost-effective maintenance concept.
- The current trend of hiving-off maintenance staff, outsourcing maintenance or integrating maintenance into production functions often result in the degradation of knowledge base and maintenance quality, and thus affecting the plant safety requirements. It is therefore important that plant life cycle communication and safety criteria be considered before outsourcing maintenance or reducing maintenance staff.

BP Texas refinery example

Texas City Refinery of BP is the third largest oil refinery in the United States. It has an input capacity of 437,000 barrels per day (18,354,000 gallons or 69,477,448

litres) as of January 2005. During start up of the isomerization unit on Wednesday March 23, 2005 following a temporary outage, an explosion and fire occurred which killed 15 and harmed over 170 people at the Texas City refinery. It was one of the most serious industrial accidents in the US in the preceding two decades. The accident was investigated by US Chemical Safety Board compiled the details of the accident in a report released on December 2005 (Chemical Safety Board, 2005).

According to the report, actions taken or not taken led to overfilling the raffinate splitter with liquid, overheating of the liquid and the subsequent over pressurisation and pressure relief. Hydrocarbon flow to the blow down drum and stack overwhelmed it, resulting in liquids carrying over out of the top of the stack, flowing down the stack, accumulating on the ground, causing a vapour cloud, which was ignited by an unknown source (probably a vehicle engine or unshielded wiring in nearby office trailers).

Accident description

The US Chemical Safety Board (CSB) investigating the incident found that operators had started up the raffinate splitter tower (which separates light and heavy gasoline components) of the isomerization unit (which increases the octane rating of gasoline) and began filling it with hydrocarbon fluid (*i.e.*, gasoline components) without beginning timely discharge of product.

The operators started the tower while ignoring open maintenance orders on the tower's instrumentation system. The design of the level indicator meant that it only read a length of 3 m; it didn't register anything above that. In addition, the design was such that any level above 3 m could show on the screen as a drop in level. There was a secondary alarm which should have gone off if liquid exceeded 2.5m; however no-one heard it; the alarm was reported as damaged before the accident and there were no records of it being fixed, so out of the two alarms, one was deactivated and the back-up never worked in the first place.

Once the lack of drawdown from the tower was recognized, operators opened the discharge valve. This worsened the problem because the hot discharges passed through a heat exchanger that pre-warmed incoming fluids. The resulting increase in temperature caused the formation of a bubble of vapour at the bottom of the raffinate tower that was already overly full and overheated. The tower burped the vapour bubble and the liquid above the bubble into the overhead relief tube of the tower.

Conclusions

According to the accident investigation team, there were four critical factors without which the incident would not have happened or would have been of significantly lower impact. These were, loss of containment of events, raffinate splitter start up procedures and application of knowledge and skills, control of work and trailer sitting, and design and engineering of the blow down stack. The investigation identified numerous failings in equipment (*e.g.* alarm system), risk management, staff management, working culture at the site, maintenance and inspection and general health and safety assessments.

This recent incident underlines the important role of maintenance in ensuring plant safety. The lack of maintenance on equipment instrumentation contributed

heavily to the accident and the consequences that followed. We therefore hypothesize that maintenance for safety is an indispensable function for all industrial systems.

22.3.4 Human Errors in Maintenance

Human errors occur for various reasons and different actions are needed to prevent or avoid the different sorts of error. Kletz (1992) classifies human errors in the following categories:

- Errors due to slip or momentary lapse of attention;
- Errors due to poor training or instructions;
- Errors due to lack of mental or physical ability (mismatch between personal abilities and the situation);
- Errors due to lack of motivation; and
- Errors made by managers due to lack of better designs, training, *etc.*

As with most types of work, the scope for human error in maintenance operations is vast. This can range from becoming distracted and forgetting important checks to knowingly deviating from a permit to work procedure in order to save time or to get the job done in unexpected circumstances. Some types of human error can be so frequent that they almost become the accepted custom and practice. For example, fitters may get into the habit of omitting final checks during a routine maintenance procedure. Other forms of human error may only occur rarely during exceptional circumstances. For example, crews may mis-diagnose the cause of a failure. In all cases, poor repairs can increase the amount of breakdowns, which in turn can increase the risks associated with equipment failure and personal accidents.

A maintenance operator who is motivated, well trained, under no time pressure, given the correct information, and working with equipment that has been designed to be maintenance friendly, will likely complete all specified maintenance work to a high standard. However, the more these requirements are not met, the less likely it becomes that the maintenance work will receive the desired attention and short cuts in work methods become increasingly probable. As a result, equipment can become poorly maintained causing reduced reliability or direct damage to the plant. In turn, these consequences can increase the safety risk to the maintenance operator and to other employees and the public. There are therefore a number of factors which influence the behaviour of maintenance crews and the likelihood of human error and these are classified by Manson (2003) into three types:

1. Slips and lapses: for example, a maintainer may be distracted or lose concentration and inadvertently undo the wrong hydraulic hose. As he knew what should have been done, there is little advantage in further training. If the consequences of such an error are significant then the most effective action would be to eliminate the possibility of this happening by some form of design. Interlocks or fittings that can only fit one way can physically prevent this type of error.

2. Mistakes: if a rule or work procedure has been forgotten, or never fully understood, then a maintainer could make a wrong decision. In the above example, the maintainer knew what he wanted to achieve but failed to achieve it. With this general type of error, the maintainer makes a mistake and chooses a wrong action. Training is obviously an important issue for reducing this type of error.
3. Violations: these are intentional deviations from maintenance procedures and are the most difficult area of human error. Such decisions can involve a range of issues such as the perceived advantages to the individual from a short cut, the risks of damage to plant and equipment if the work is not done, the likelihood that the maintainer will be subsequently identified; and the time allocated to the job in relation to the time the job takes to fully adhere to the approved procedure.

There will therefore be a range of factors which influence the likelihood of maintenance rule violations. These can be divided into those which directly motivate the maintenance crew/individual to break agreed rules/procedures (termed direct motives) and supplementary factors which increase, or reduce, the probability of any individual deciding to commit a violation (termed behaviour modifiers) (Manson, 2003). For example, avoiding heavy physical work may be a direct motive for neglecting a maintenance task; however, a lack of effective supervision would be a behaviour modifier that increases the probability that the violation would occur as the chances of him being detected would be low.

22.3.5 Accident Causation Theories vs Maintenance

Several theories and models are used in the literature to explain how accidents happen. These theories try to illustrate how potential safety threats (therefore referred to as hazards) are translated into injury, loss of life and/or destruction of property (therefore referred to as an accident). Using these theories, we can identify some relationships between maintenance and safety, and moreover, on how maintenance can impact plant safety.

The *Domino Theory* developed in 1931 by Heinrich suggests that one event leads to another, then to another and so on, culminating in an accident (Heinrich *et al.* 1980). In the 1920s, Heinrich studied and classified the records of 75,000 industrial accidents and concluded that 88% of industrial accidents were caused by unsafe acts of people, 10% of industrial accidents were caused by unsafe conditions, and 2% of industrial accidents were unavoidable (acts of God). Subsequent development of this theory identifies immediate causes of accidents and the contributing causes to the accident (Tania, 2003). Immediate causes involve unsafe acts, unsafe conditions and the acts of God. Contributing causes include the safety management performance, mental condition of the worker and the physical condition of the worker. The theory predicted that removal of the central factors in an accident chain, the unsafe acts and hazardous conditions, would negate the action of preceding dominos – social environment of work, and negative character traits of worker – and therefore prevent the final two dominos in the causal chain, accident and injury. This theory is supported by Raouf's work on

organizational accident causation theory. He identifies the possible accidents contributing causes as unsafe acts, unsafe conditions and organizational factors (Raouf, 2007). He states that the barriers to organizational accidents are competent and trained workers, well outlined procedures, and safety condition of plant machineries. These factors suggest some links of maintenance interaction with safety.

Maintenance staff have a key role identifying and rectifying unsafe conditions in the plant. The maintenance technician can take what engineering has designed, and reduce the equipment and environmental hazards even more. Another key role of maintenance staff is to carry out their activities in a safe manner that is unlikely to cause unsafe acts. Furthermore, maintenance people are usually deeply involved in safety-related duties such as first aid, fire brigade, and disaster planning /preparation due to their extensive knowledge of the facility layout, and thus has a high impact on safety management performance. We therefore conclude that maintenance function can impact plant safety by improving unsafe conditions, avoiding unsafe acts and improving safety management performance as show in Figure 22.3.

The *Human Factor Theory* argues that any accident is due to a chain of events ultimately caused by human error (Heinrich *et al.* 1980). Human error may be caused by factors such as physical and/or psychological factor regarding the capacity of the worker, inappropriate responses and inappropriate activities. In this theory, the total load includes task responsibilities, environmental factors, internal factors, and situational factors. An extension of the human factors theory is the *Accident/Incident Theory* that added ergonomic traps and wilful decision to err to the overload conditions as a more comprehensive look at human error causes (Tania, 2003). It further stated that accidents occur due to system failure. Based on these theories, maintenance can contribute by reducing the overload due to environmental factors (*e.g.*, reduce noise level), reducing situational factors (*e.g.*, prevent oil leakage onto floor) as well as preventing the system failure.

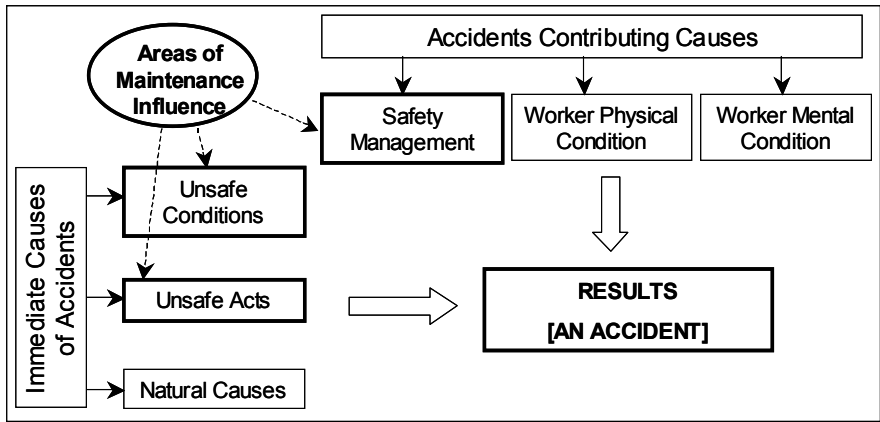


Figure 22.3. Areas of maintenance influence on immediate causes and contributing causes of accidents

In the *Epidemiological Theory* of accident causation, the key components are the pre-dispositional characteristics of the worker and the situational characteristics of the job. These work together to cause or prevent an accident. Maintenance may have a bigger potential impact on the situational characteristics of the job, but through influence on the worker, could modify a predisposition to violate operating procedures *e.g.*, override safeguards, *etc.*

The *System Theory* model states that there are three main components that interact in any job: the worker, the machine/equipment, and the environment. The likelihood of an accident is determined by how these components interact. Changes in the pattern of interaction can increase or reduce the probability of an accident occurring (Goetsch, 1999). However, the elements that interact in the manufacturing process go beyond the workers and the machines as indicated in the system theory. In combination with the workers and machines, the other elements that have a great impact on plant safety are the material being handled in the production process and the method of production. We therefore identify four elements whose interaction in the plant generates various outcomes. These four elements will be referred to as man, material, machine, and method. The four elements interact with each other in a given manufacturing environment to give the various outcomes as shown in Figure 22.4 (based on Pinjala, 2007).

Under normal conditions, the four elements interact in an anticipated way to produce products. However, any unanticipated or unexpected interactions between these elements can result in any one or a combination of events. This can be a production loss, defective or scrap products, failed or damaged equipment, an accident or even pollution to the outside environment. The rate of these incidents depends upon the configuration and level of interaction of the elements.

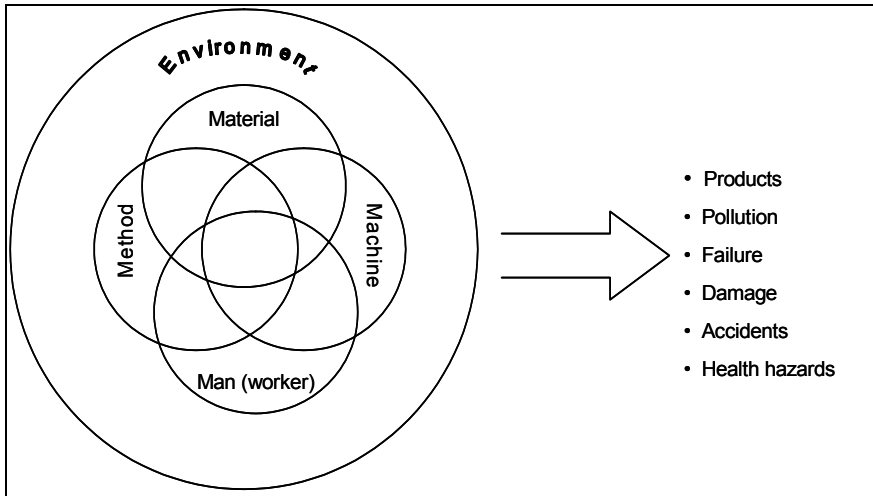


Figure 22.4. The interaction of man, machine, method and material and the possible outcomes

The concept of man, machine, material and method interactions can be extended to illustrate best the interaction between maintenance and safety. The maintenance staff interacts with machines during corrective or preventive maintenance. For the maintenance function to be successful, the correct method (in this case, the working procedures) need to be followed with the aid of correct tools and materials. If procedures (and tools) are not followed correctly during maintenance, accidents are likely to happen during maintenance work causing some casualties to maintenance staff. Furthermore, faulty maintenance may result, thereby causing the plant to fail dangerously during start-ups or operation. Maintenance may have a positive impact on equipment/machine (*e.g.*, design and installation of safety guards that cannot be disabled or removed) and environment (*e.g.*, noise reduction, control of surface temperatures of machines the worker may touch, ventilation of toxic materials, and illumination).

The theories explained above of accident causation indicate that there is a maintenance link to plant safety. It also signifies the role of maintenance workers in accident prevention to operational worker and to the fellow maintenance workers.

22.4 Maintenance Policies and Concepts vs Safety

Maintenance in the manufacturing environment can be broadly explained in terms of maintenance actions, maintenance policies and maintenance concepts. The maintenance actions, policies and concepts adopted in a certain plant have a big impact on plant safety and the safety of the maintenance work. Before we look at the safety implications of each maintenance policy and concepts, let us first see the definition of each term.

22.4.1 Definitions

Confusion does exist in both literature and practice on the meaning of maintenance actions, policies and concepts. What some call concept is a policy to others; what some call policy is a maintenance action to others. Pintelon and Van Puyvelde (2006) distinguishes these terminologies by the following definitions:

- *Maintenance actions*: the basic maintenance interventions and the elementary work carried out by a technician. It is a question of what do maintenance staff do?
- *Maintenance policy*: these are rules or set of rules describing the triggering mechanism for the different maintenance actions. It is a question of what triggers maintenance actions?
- *Maintenance concepts*: these are set of maintenance policies and actions of various types and the general decision structure in which these are planned and supported. It is a question of which maintenance decision structure is used?

Maintenance actions entail the activities taken by the technicians at operational level. These may be corrective actions or precautionary actions. Maintenance policy entails the set of rules that triggers the maintenance actions and can be classified as a tactical decision level. Some examples are failure-based maintenance (FBM), condition based maintenance (CBM), opportunity based maintenance (OBM), *etc.* Finally, the maintenance concept entails the general decision structure for both maintenance actions and policies and can be classified as a strategic decision element. Some examples are reliability centred maintenance (RCM), total productive maintenance (TPM), business centred maintenance (BCM) among others.

22.4.2 Maintenance Actions

Maintenance actions can either be corrective (CM) or precautionary (PM):

- *Corrective actions* are repair or restore actions taken after a breakdown or a loss of function. Corrective actions are difficult to predict as equipment failure behaviour is stochastic and breakdowns are unforeseen. For example, a ruptured pipe, a stuck bearing or broken gear teeth will need a corrective action. Corrective maintenance has the highest interaction with plant safety and is a source of many safety hazards and accidents in industry. The source of safety hazards or accidents may first be due to the failure itself. In this case, the extent of the safety hazard is dependent on the criticality of the equipment or the component and the extent of the failure. For example, failure on pressurized equipment like a boiler has high consequences for safety. Due to the nature of corrective maintenance jobs, there is barely no time to prepare or follow the procedures correctly. Combined with the pressure to restore production, corrective maintenance may lead to accidents during maintenance or after faulty maintenance.

- *Precautionary actions*, often referred to as preventive actions, are actions mainly aimed at diminishing the failure probability and or the failure effect. These preventive actions are easier to plan because they rely on fixed time schedules or

prediction of stochastic behaviour. Examples of precautionary actions are lubrication, oil and filter change, periodic bearings change, inspections change, vibration monitoring among others. These kinds of maintenance actions give the best approach to reducing and containing maintenance related accidents. The precautionary actions may be predictive, preventive or proactive and aims at notice failure before it actually happens. These precautionary actions support the common wisdom that prevention is better than cure. Due to adequate time to prepare these actions, maintenance procedures are more likely to be followed correctly, thereby reducing chances of incidents. Implementation of precautionary maintenance actions helps to mitigate accidents or incidents related to equipment and workers, and leads to zero failures and zero accidents.

22.4.3 Maintenance Policies

Maintenance policies outline the rules for triggering maintenance actions and are therefore important tactical level decisions. Several types of maintenance policies can be considered to trigger, in one way or another, either precautionary or corrective maintenance interventions. These policies are mainly failure-based maintenance (FBM), time/used-based maintenance (TBM/UBM), condition-based maintenance (CBM), design-out maintenance (DOM) and opportunity-based maintenance (OBM). Maintenance policies are either reactive, preventive, predictive, proactive or passive. It is worth noting that the formation of maintenance policies are based not solely on technical considerations but rather on techno-economic considerations. The kind of policies adopted for the plant or for a specific equipment has great impact on maintenance activities, productivity, and plant safety.

FBM is a purely reactive policy. Maintenance is carried out only after breakdown. The main aspects considered by the industry are the cost of CM vs costs of alternative PM, risks for and consequences of secondary damage and potential safety hazards. Since no planning is possible, unforeseen breakdowns disrupt production and spares and manpower should be kept available to solve the problem as soon as it occurs. This method may be appropriate for plants like glass ovens, where cooling down the oven for preventive intervention takes too much time – (several) days – and a lot of energy to heat it again. However, reactive maintenance is a recipe for safety hazards as some past statistics may tell us. A recent survey shows that 60% of all safety incidents occurred when a maintenance job was executed as reactive. The data was collected from many industries where pulp and paper industry represented 36 of all respondents (IDCON, 2007). Another study done in paper companies concluded that it was 28% more likely to have an incident when maintenance work was reactive vs planned and scheduled before execution (IDCON, 2007). It makes sense that there is a strong correlation between safety incidents, injuries and reactive maintenance. In a reactive situation, you might not take time you should to plan and think before you take action. The urgency also calls out the so common hero in maintenance craftsmen and they take risks that they should not take.

UBM/TBM are preventive maintenance policies where maintenance is carried out at specified time intervals. For UBM, intervals are measured in working hours while in TBM intervals are in calendar days. In between PM actions, CM actions

can be carried out when needed. Either TBM or UBM is applied if the CM cost is higher than PM cost, or if it is necessary because of criticality due to the existence of bottleneck installation or safety hazards issues. Also, in case of increasing failure behaviour, like for example wear-out phenomena, TBM and UBM policies are appropriate. Many interval optimization models are available and they try to balance PM and CM costs. However, TBM or UBM policies are unable to foresee failure and are therefore unable to reduce the failure probability. Hence, safety hazards can still be realised. This problem is addressed through CBM policy, if there exists a measurable condition, which can signal the probability of a failure.

Initially, CBM was mainly applied for those situations where the investment in condition monitoring equipment was justified because of high risks, like aviation or nuclear power regeneration. With the reduction in implementation costs, the predictive techniques are generally accepted to maintain all types of installations. Furthermore, CBM catches the attention of practitioners due to the potential savings in spare parts replacements thanks to accurate and timely forecasts on demand. In turn, this may enable better spare parts management through coordinated logistics support. This predictive policy is one of the best as far as plant safety is concerned. It is able to mitigate failures long before they occur and give maintenance staff some adequate time to prepare PM actions. The main challenge with this policy lies in finding and applying a suitable CBM technique for each scenario. For example, the analysis of the output of some measurement equipment, such as advanced vibration monitoring equipment, asks for a lot of experience and is often work for experts. But there are also simpler techniques such as infra-red measurement and oil analysis suitable in other contexts. At the other extreme, predictive techniques can also be rather simple, as is the case of checklists. Although fairly low-level CBM these checklists, together with human senses (visual inspections, detection of “strange” noises in rotating equipment, *etc.*), can detect a lot of potential problems and initiate PM actions before the situation deteriorates to a breakdown. With the development and improvement of BITE (built in test equipments), CBM is getting better and better.

While FBM, TBM, UBM and CBM accept and seize the physical assets which they intend to maintain as given, there are more proactive maintenance actions and policies, which look at the possible changes or safety measures needed to avoid maintenance in the first place. This proactive policy is referred as DOM. This policy implies that maintenance is proactively involved at earlier stages of the product life cycle to solve potential problems in relation to maintenance. Ideally, DOM policies intend to avoid maintenance completely throughout the operating life of installations, though this may not be realistic. Then the basic idea turns out to include a diverse set of maintenance requirements at the early stages of equipment design. As a consequence, equipment modifications are geared either to increasing reliability by rising the mean-time-between-failures (MTBF) or to increasing the maintainability by decreasing the mean-time-to-repair (MTTR). *Per se*, DOM aims to improve the equipment availability and safety. Often DOM projects are used to support efforts to increase occupational safety as well as production capacity.

A rather passive but considerably important maintenance policy that needs to be mentioned is OBM. OBM is applied to non-critical components with a relatively long lifetime. For these components no separate maintenance programs

are developed; maintenance take place if an opportunity arises because there is maintenance intervention for another component of that machine. As previously stated, this policy may not be applied in installations that pose safety hazards.

22.4.4 Maintenance Concepts

The holistic view of a maintenance program suggests that an adequate mix of maintenance actions and policies needs to be selected and fine-tuned in order to improve uptime, extend the total life cycle of physical assets and assure safe working conditions, while considering limiting maintenance budgets and environmental legislations. Therefore, a maintenance concept for each installation is necessary to plan, control and improve the various maintenance actions and policies applied. As a matter of fact, maintenance concepts need to be formulated considering the physical characteristics and the context installations operate. Not surprisingly, as system complexity increases and maintenance requirements become more demanding, maintenance concepts also advocate different levels of complexity. A maintenance concept is important because in the long term it may even become a philosophy to perform maintenance. Maintenance concepts also determine the business philosophy concerning maintenance, and they are needed to manage the complexity of maintenance *per se*. No doubt, the maintenance concept adopted has a big influence on maintenance-safety interactions in a plant.

Literature provides us with various concepts, that have been developed through a combination of theoretical insights and practical experiences. Typical examples, and perhaps the most important ones, are Total Productive Maintenance (TPM), Reliability-Centred Maintenance (RCM) and Life Cycle Costing (LCC) approaches. Unmistakably, these concepts enjoy several advantages as well as some specific shortcomings. Some are supported by a number of consultants who make profits out of them. In the same way, more and more companies are searching for their own customised concepts. The main challenge lies on choosing and implementing the best concept in a given context. There is no short and straightforward answer to the question “what concept is best for us?”. The right answer to the question is determined by the context, with its complex interaction of technology, business, organization, and, indeed the plant safety.

We shall look at some of these concepts and see how they impact plant safety;

22.4.4.1 *Quick and Dirty Decision Charts (Q&D)*

A Q&D decision chart is a decision diagram with questions on failure patterns and repair behaviours of the equipment, on business contexts, on maintenance capabilities, on cost structure and so forth. Answering the questions for a given installation, the user proceeds through the branches of the diagram, and the process stops with the recommendation of the most appropriate policy for the installation on-hand. The Q&D approach allows for a relatively quick determination of the likely most advantageous maintenance policy. It ensures a consistent decision making for all installations. Although some Q&D decision charts are available from the literature, *e.g.*, Pintelon (2000), most companies adopting this approach prefer to draw up their own charts, which incorporate their insights based experience and knowledge in the decision process. This approach however has the

drawback of being rough (dirty). The questions are usually put in the basic yes/no format, limiting the answering possibilities. Moreover, answering the questions is usually done on a subjective basis; for example the question whether a given action or policy is feasible is answered based on experience rather than on a sound feasibility study. From the safety perspective, the Q&D approach can be used to identify the appropriate maintenance policy especially for a critical equipment that poses high safety hazard. If thoroughly applied, the approach can indicate the maintenance policy for each piece of equipment and therefore support mitigation of maintenance related incidents. However, the same drawback of being quick and dirty applies to safety considerations in a plant. If some safety aspects are overlooked in decision chart, it may be disastrous for the plant.

22.4.4.2 Life Cycle Costing (LCC) Approaches

Life cycle costing (LCC) is a methodology to calculate and to follow up overall cost of a system from inception to disposal (that is, during the entire course of its life). First, there is cost iceberg structure as launched in 1981 and later developed by Blanchard (1992). The iceberg warns that it is not only the initial purchase cost of an installation that is important; there are other cost that are relevant too, which are mostly ignored in investment decision making. But indirectly the relevant long run costs such as operational expenses, training cost, maintenance costs, spares inventory costs, *etc.*, are at least of the same order of magnitude. The cheapest machine is not always the cheapest one in terms of maintenance and operation.

LCC also refers to the principle that the further one gets in the design or construction cycle of equipment, the more costly it will be to make modifications; think for example about DOM. It draws attention to the fact that many of the costs that will be needed to operate and maintain equipment are fixed at the design phase. It is of the utmost importance to consider all aspects of whole intended life of the equipment from the design phase on. Maintenance should therefore be taken into account from the very first moment of designing a machine or system.

The LCC approach implies a synthesis of costing analysis and engineering design principles that must satisfy life cycle requirements at minimum cost with design decisions being based on total cost of ownership (TCO) principles. However, much emphasis is not put on equipment operation safety or to plant safety in general. It is also a fact that the equipment with a minimal life cycle cost is not necessarily the safest. In the process of minimising the cost, equipment safety may be compromised, leading to even higher costs. However, with the development and application of some of the LCC approaches like terotechnology (developed in UK in the 1970s) (Parkes and Jardine, 1970), design issues of the equipment's maintainability and reliability are taken into consideration. Terotechnology is concerned with the specification and design for reliability and maintainability of physical assets and takes into account the processes of installation, commissioning, operation, maintenance, modification and replacement.

Consideration of safety related cost could be more value adding to LCC approach. Lack of safety may have a high price tag due to loss of life "cost", property destruction cost, environmental pollution cost, insurance cost, loss of production cost, stoppage or shutdown cost. However, some intangible aspects like bad reputation and loss of goodwill "cost" after an accident cannot be captured by

LCC. Consideration of these costs against equipment reliability costs would be very interesting in LCC approach.

22.4.4.3 Total Productive Maintenance (TPM)

TPM is based on Productive Maintenance, which was introduced in the 1950s at General Electric Cooperation. Later on it was further developed in Japan and re-imported in the West (Takahashi and Takashi 1990). TPM goes beyond a maintenance concept and is sometimes translated as Total Productive Manufacturing. TPM involves total participation at all levels of the organization. It aims at maximizing equipment effectiveness and establishing a thorough system of preventive maintenance. TPM fits entirely with the TQM philosophy and the JIT approach. The TPM toolbox consists of various techniques, some universal ones such as 6 sigma, Pareto or ABC analysis, Ishikawa or fishbone diagrams, *etc.* In addition, other more specific concepts and techniques such as SMED, poke yoke, jidoka, OEE, and the 5S. The overall equipment effectiveness (OEE) is a powerful tool to measure the effective use of production capacity. The strength of the concept is the integration of production, maintenance and quality issues into what is called the “six big losses” of useful capacity. On the other hand, the 5S form one of the basic principles of TPM: *Seiri* (or sorting out), *Seiton* (or systematic arrangement), *Seiso* (or Spic and span), *Seikutsu* (or standardizing) and *Shitsuku* (or self-discipline).

Nakajima, commonly accepted as the father of TPM, describes the concept in the following five points (Nakajima, 1989): (1) aims at getting the most efficient use of equipment (improve overall effectiveness), (2) it establishes a complete productive maintenance program encompassing maintenance prevention, preventive maintenance, and improvement related maintenance for entire life cycle of the equipment, (3) it is implemented on a team basis and it requires the participation of equipment operators, and maintenance technicians, (4) it involves every employee from top management to the workers at the shop floor, and (5) it promotes and implements productive maintenance based on autonomous small-group activities.

TPM seeks to go beyond preventive maintenance towards prevention of maintenance by eliminating maintenance related problem, improving plant reliability and improving plant's design. By achieving these objectives, the plant can run with zero defects, zero breakdowns and zero accidents. Since it promotes teamwork and cooperation of all employees, standard operating procedures can easily be followed thereby promoting more plant safety. Less corrective or accidental maintenance translates to less/no accidents during maintenance and less/no accidents due to lack of maintenance. TPM concept seeks to improve productivity (and thus profitability), by improving equipment effectiveness through quality maintenance. The recent TPM has explicitly incorporated safety and environmental management.

22.4.4.4 Reliability Centered Maintenance (RCM)

RCM originates from the 1960s in North American aviation industry. Later on, it was adopted by military aviation, and afterwards it was only implemented at high-risk industrial plant such as. nuclear power plants. Now it can be found in industry at large. Well-known are the books by Nowlan and Heap (1978) Anderson and

Nari (1990) and Moubray (1997) who contributed to the adoption of RCM in industry. Note that today many versions of RCM are around, streamlined RCM being one of the more popular ones. However, the Society for Automotive Engineers (SAE) holds the RCM definition that is generally accepted. SAE puts forward the following basic questions to be solved by any RCM implementation; if any of those are omitted, the method is incorrectly being referred as a RCM. To answer these seven questions a clear step-by-step procedure exists and decision charts and forms are available:

1. What are the functions and associated performance standards of asset in its present operating context?
2. How can it fail to fulfil its functions (functional failures)?
3. What causes each failure (failure modes)?
4. What happens when each failure occurs? (failure effects)
5. In what way does each failure matters? (failure consequences)
6. What should be done to predict or prevent each failure (proactive tasks and task intervals)?; and
7. What should be done if a suitable proactive task cannot be found (default actions)?

RCM is undeniably a valuable maintenance concept. It takes into account system functionality, and not just the equipment itself. The focus is on reliability rather than maintainability and availability. Safety and environmental integrity are considered more important than costs. Applying RCM helps to increase the assets' lifetime and to establish a more efficient and effective maintenance. Its structured approach fits in the knowledge management philosophy: reduced human error, more and better historical data and analysis, exploitation of expert knowledge and so forth. Some authors (Waeyenbergh and Pintelon 2002) argue that this approach is justifiable in aircraft industries and in high risk industries, but it is often too expensive in general industries where maintenance is an economic rather than a reliability problem.

Though expensive and tedious, RCM offer the best safety oriented approach of all the other maintenance concepts. The issue of faulty maintenance or failure due to lack of maintenance is not meant to arise in RCM. To ensure plant/system reliability, maintenance is carried out accurately, with respect to laid down procedures, and without undue pressure from operations. This also reduces the chances of accidents during maintenance. No wonder it is the most recommended concept for high risk systems to ensure maximum safety. With the use of tools like FMEA (failure modes and effect analysis), FTA (fault tree analysis), ETA (event tree analysis), RCA (root cause analysis) and HAZOP, RCM is able to get to the root cause of failures and eliminate them. RCM therefore offers the best safety-oriented approach to maintenance and to the plant.

22.5 Maintenance Safety and Accident Prevention

As stated by Levitt, accident or hazard control and prevention are actions directed toward recognizing, evaluating, and eliminating (or reducing) the risk of hazards

emanating from human errors and from the situational and environmental aspects of the workplace (Levitt, 1997). This process can occur organization-wide, department-wide, by machine, or even by individual component. Human errors that could potentially cause an accident are called unsafe acts, and may be defined as being human actions that depart from hazard control or job procedures to which the person has been trained or otherwise informed, which causes unnecessary exposure of a person to a hazard or hazards. Situational and environmental hazards may enter the workplace from many sources: (1) purchased parts or materials, and how they are produced, packaged, and labelled; (2) engineers responsible for tool and machine design, their placement in the workplace, and provisions for adequate warnings and machine guards; and (3) those responsible for maintaining shop equipment, machinery, and tools. This third source, maintenance activities, leads to a fundamental tenet of safety management that no hazard control program can succeed if housekeeping and maintenance are not seen as integral parts.

Seen in this perspective, maintenance is definitely a major resource to abate and mitigate safety problems. Maintenance workers with proper management and work instructions/time can identify hazards, repair potential safety problems for other workers, and be advocates for increased safety. They can do this during repairs (corrective maintenance) and especially during preventive maintenance (PM) which involves orderly, uniform, continuous and scheduled action to prevent breakdowns, prolong the useful life of equipment, assure quality output of the equipment, and assure safe equipment operations and maintenance in the future.

22.5.1 Methods of Accidents and Hazards Avoidance in Maintenance

There are four accepted approaches to industrial hazard avoidance (Batson *et al.* 1999):

- Analytical approach;
- Engineering approach;
- Enforcement approach; and
- Psychological approach.

The enforcement and psychological approaches focus on prevention of unsafe acts and have much correlation with human error. While these factors are important for plant safety and are applicable for all plant workers, the analytical and engineering approaches provide more insight into the role of maintenance in plant safety and are discussed in more details below.

22.5.2 Analytical Approach

The *analytical approach* deals with hazards by studying their mechanisms, collecting and analyzing historical data on accidents and incidents where the hazard was a causal factor, computing probabilities of events leading up to and including accidents, conducting epidemiological and toxicological studies, and weighing cost/benefit of hazard elimination alternatives. Such computational approaches would appeal to maintenance engineers who could work with safety engineers, computer scientists, and/or statisticians to carry out meaningful

analytical studies. Among the most popular analytical approaches to hazard prevention are Goetsch (1999), Pintelon (2006), Pintelon *et al.* (2000):

- Accident Root Cause Analysis (RCA);
- Failure Modes and Effects Analysis (FMEA);
- Fault Tree Analysis (FTA);
- Hazard and Operability Analysis (HAZOP); and
- Human Error analysis (HEA).

Accident RCA is the most widely practiced of the above three approaches. After an accident occurs, almost every plant conducts an accident cause analysis or has one performed by an outside expert. Certainly, accident cause analysis can provide information to the maintenance department on how they can change repair procedures/schedules, better label parts, pipes, *etc.*, better instruct operators, or design and install safeguards – all with prevention of similar accidents as the goal. Maintenance may also be asked to work with equipment engineers on design changes, or operational supervisors on procedural changes that will serve to protect the worker. Should the accident have occurred during maintenance, then obviously maintenance management should be involved instead of operational management.

Failure Modes and Effects Analysis (FMEA) looks at a product in operation, or the manufacturing process for the product, and identifies failure modes – what could fail in the equipment. Hence, it is not directly a safety analysis method, but indirectly it does identify effects of each failure mode and among these may be conditions that could lead to an industrial injury or illness. Maintenance can make use of FMEA even before an accident. Every component of equipment has some feasible mechanism for eventual failure that can be identified. The FMEA can direct attention to critical components that should be set up on a PM policy, which permits parts to be inspected and replaced before failure.

Fault Tree Analysis (FTA) is a system safety tool for modelling chains of cause and effect leading to some undesirable event, such as an accident. All procedural and equipment-related causes are considered, so it is more flexible than FMEA and has been used for safety analysis and equipment/procedure design in many industries, including defence, space, and nuclear power generation. The chain of cause and effect is modelled as a Boolean Tree, with alternating levels of and-gates and or-gates describing the logic of the focal (head) event. After the logic of the potential causal chains is defined, probabilities are adjoined to the tree – one for each event in the tree – and the laws of probability are used to calculate the probability of the head event. Then engineering or procedural controls are proposed, their impact on the probabilities in the tree estimated, and the probability of the head event recalculated with these preventive actions “in place.” Many alternatives can be tested, and along with their total costs, provide a cost-benefit analysis for engineers and company management to choose the action that best fits the company situation.

22.5.3 The Engineering Approach

The Engineering Approach is effective against many pieces of equipments and environmental hazards. It is considered highly preferable when dealing with health

and safety hazards in the workplace. The engineering approach presents three lines of defence against safety hazards:

1. Engineering controls;
2. Safety procedures for maintenance work; and
3. Personal protective equipment (PPE).

22.5.3.1 Engineering Controls

Engineering controls arise from previous experience with similar equipment, company, industry, or government-enforced standards, or practical experience with the equipment in question. Maintenance can relate accident information back to equipment engineers in detail, can assure engineers follow the standards, or can redesign and modify equipment already in place. Fail-safe principles of design, and equipment shut-off, also are examples of engineering controls.

Engineering controls also include protective systems used to protect operating plant against over-pressurisation and release of toxic materials, and process control instruments, which are linked to plant safety. The choice and specification of any protective system requires a careful study of both the events it is intended to mitigate or avert and the extent to which such protection is provided in the basic design (King, 1990). General codification is thus difficult. Manufacturers of this protective equipment have provided lists of basic and special preventive features, with guidance on their applications, *e.g.*, fire and explosion protective equipment. Another example is the American Petroleum Institute's (API, 1976) recommendations on the choice, design and installation of over-pressure relief systems for oil refineries which has wide application in the whole process industry. The specification of protective systems is best done in conjunction with a HAZOP study (King, 1990). Good preventive maintenance plays a major role in ensuring that hazard controls stay in place and remain effective as well as prevent new hazards from arising due to equipment malfunction.

22.5.3.2 Safety Procedures for Maintenance Work

In spite of maintenance importance to the plant, many plant accidents have occurred during or following maintenance because of misunderstanding and neglect of essential precautions when plant was handed over from production to maintenance workers and *vice versa*. The maintenance workers may be company employees or may be employed by an outside contractor. The possibilities of misunderstandings between operating and maintaining personnel are aggravated by shift work and by the use of outside contractors. Work procedures are therefore important before any maintenance work can begin, then during the maintenance process and after maintenance especially when handing the equipment back to production. Careful planning of procedures is important for both small and big jobs or during routine or emergency jobs. Among the important procedures demanded for safe maintenance practices as stipulated by American Petroleum Industries (API, 2007) are:

- Orders for maintenance work should be authorized in writing with a description of work to be performed. This is often referred to as work permit (King, 1990).

- Every maintenance job plan should include specific instructions for the execution of the job such as the estimated man-hours, craft sequences, reference to applicable drawings and sketches, material required, equipment (including fire and safety equipments) to be provided, reference to standard for a particular job which may differ from standard practices, priority for the job among other instructions.
- Maintenance job orders must be dispatched well in advance of the start of work so that:
 - The field maintenance supervisor will have time to study the job and establish proper liaison with operating and fore safety personnel before the work is started;
 - Pertinent standard and special practice instructions may be reviewed beforehand;
 - On-the-job safety meetings may be held as needed to brief the personnel on special hazards and techniques;
 - Adequate facilities for the transportation of men and delivery of materials, including tools and special material required, may be scheduled; and
 - Other departments concerned, including fire and safety department, may be notified in sufficient time to provide the necessary permits and equipments.
- The execution of the job should be closely followed so the planned performance will produce the expected results. It is well to observe whether the job methods utilized are safe and efficient.
- As the maintenance job commences, careful attention to instruction and duties, use of right tools and proper use of protective equipments is needed to minimize possibilities of accidents and injuries.
- Careful attention should be given to hot lines and equipments, rotating and reciprocating equipments, furnace gases and vapours, electric connections to the equipment being worked on, oil spills, open trenches or sewers, electric welding arcs, congested pathways, sharp objects, inadequate ventilation among other potential hazards.
- For the equipments being worked on, tagging and locking is imperative so that operators may not run the equipment when maintenance is working on it.
- When the maintenance work is completed and equipment is ready for production, operating and maintenance supervisor should inspect the equipment together and assure that it is safe for operation.

The laid down procedures may vary from one industry to another or from one piece of equipment to another.

22.5.3.3 Personal Protective Equipment

Personal protective clothing and/or equipment (abbreviated PPC/E) is needed against particular hazards of the working environment. This is particularly important for maintenance workers who work in potentially dangerous environments or with potentially dangerous equipment. The use of PPE in maintenance goes beyond accident prevention to protection against occupational diseases.

Depending on the working environment of maintenance, several parts of the body or the whole body may need protection. Among the most important protective equipments are (King, 1990);

- Hand protection equipments (*e.g.*, gloves);
- Head protection equipments (*e.g.*, head helmets or welders helmets);
- Foot protection equipments (*e.g.*, safety boots/shoes);
- Eye protection equipments (*e.g.*, safety goggles, spectacles);
- Hearing protection (*e.g.*, ear muffs or ear plugs);
- Respiratory protection equipments (*e.g.*, respirators, breathing apparatus); and
- Body protection (*e.g.*, hot working clothing, clean-working clothing and general aprons).

The protective clothing and equipment should be available to employees where needed. Workers would also require some training on when and how best these PPE should be used. If correctly used, the PPE prevents injuries or reduces the severity of the injuries should the accident happen.

22.5.4 Safety Culture

As seen in Section 22.1 above, the variety of risks associated with industries can be managed in different ways, for instance through rules and procedures, training, supervision, use of PPE, engineering controls and risk assessment. However, these risk mitigation methods may not be enough to prevent accidents without change of attitude, participation of every employee, support from management, *etc.* All these aspects combined define the safety culture of a company. This involves creating a culture within an organization where everyone is personally involved in ensuring safety and where the values of safety are evident in every activity from general company policy and philosophies to the actions of a front line operator (Hudson, 1999). Though safety culture is an important concept, a single definition has not been agreed on. A definition of safety culture by Health and Safety Executive states that (HSE, 1999):

“The safety culture of an organization is the product of individual and group values, attitudes, perceptions, competencies, and patterns of behaviour that determine the commitment to, and the style and proficiency of an organisation’s health and safety management. Organizations with a positive safety culture are characterised by communications founded by mutual trust, by shared perceptions of the importance of safety, and by confidence in the efficacy of preventive measures.”

Without a positive safety culture and climate, there would be resistance to safety schemes and programs being implemented, possibly dooming them to failure from the outset. Lack of safety culture may explain the initial resistance to safety initiatives and lack of staying power in these initiatives to bring about a permanent change or some degree of change (Darby *et al.* 2005). Due to the risky nature of a maintenance job, a positive safety culture is imperative. Promotion of a positive safety culture is therefore considered a viable way of managing risk and an effective way of accident avoidance. This goes beyond the maintenance department to the production department and to the whole industry.

Safety culture explains how safety is regarded as a priority within an organization. It may be reflected in decision and policies of the organization and filters down through these into every aspect of operational performance. It governs the conduct and behaviour of every employee and promotes safety consciousness. Several factors have been identified as supporting development of a positive safety culture within various industries. Key amongst them are management, immediate supervisors, individual and behavioural factors, reporting systems, rules and procedures, communication and organizational subcultures and subcontractors (Darby *et al.* 2005).

Though safety culture is a potentially valuable concept, it is rather a vague concept. It is a perception or attitude to safety and it cannot be easily measured or quantified. It cannot be directly managed but may be influenced by some managerial initiatives.

22.5.5 Safety Legislations

Since the industrial revolution, the amount of legislation passed and the number of subsequent regulations concerning workplace health and safety have increased remarkably. Of all these legislations, by far the most significant has been Occupational Safety and Health Act of 1970, called the OSHA Act (King, 1990). The Occupational and Safety Health Act was created to protect worker and workplace safety. Its main aim was to ensure that employers provide their workers with an environment free from dangers to their safety and health, such as exposure to toxic chemicals, excessive noise levels, mechanical dangers, heat or cold stress, or unsanitary conditions. The OSHA act was enacted through the Occupational Safety and Health Administration (OSHA) agency of the US Department of Labour. The mission of the agency is to prevent work-related injuries, illnesses, and deaths by issuing and enforcing rules (called standards) for workplace safety and health. According to US labour department, the mission and purpose of OSHA can be summarised as follows (Goetsch, 1999):

- Encourage employers and employees to reduce workplace hazards;
- Implement new health and safety programs;
- Improve existing health and safety programs;
- Encourage research that will lead to innovative ways of dealing with workplace health and safety problems;
- Establish the rights of employers regarding the improvement of workplace health and safety;
- Establish the rights of employees regarding the improvement of workplace health and safety;
- Monitor job related illnesses and injuries through a system of reporting and record-keeping;
- Establish training programs to increase the number of health and safety professionals and to continually improve their competence;
- Establish mandatory workplace health and safety standards and enforce those standards;
- Provide for the development and approval of state-level workplace health and safety programs; and

- Monitor, analyze and evaluate state-level health and safety programs.

Much of the debate about OSHA regulations and enforcement policies revolves around the cost of regulations and enforcement, vs the actual benefit in reduced worker injury, illness and death. A 1995 study of several OSHA standards by the Office of Technology Assessment (OTA) found that regulated industries as well as OSHA typically overestimate the expected cost of proposed OSHA standards (OTA, 1995).

Another organization that is actively involved in legislations for occupational safety and health is the International Labour Organization (ILO). It is an agency for the United Nations that promotes opportunities for people to obtain decent and productive work, in conditions of freedom, equity, security and human dignity. However, its mandate goes beyond occupational safety and seeks to promote employment creation, strengthen fundamental principles and rights at work, improve social protection, and promote social dialogue as well as provide relevant information, training and technical assistance.

Besides the abovementioned organizations, there are many more organizations that concern themselves with occupational safety and health. Many of these organizations have informative websites. A good example of such a website is *osha.europa.eu*, the website of the European Agency for Safety and Health at Work. The agency, founded in 1996, states its mission as making Europe's workplaces safer, healthier and more productive, and in particular promoting an effective workplace prevention culture. On the website interesting information is provided (brochures, guidelines, good practice examples, tools and checklists, *etc.*) as well as links to the national websites for safety and health at work of the different European member countries and links to international sites such as OSHA and ILO and similar organizations in Australia, Canada, Japan, Korea and the USA. As an illustration of the practical information offered on national websites, we refer to the Belgian governmental organization for safety and health at work (responsibility of the *Federale Overheidsdienst Werkgelegenheid, Arbeid en Sociaal Overleg, Welzijn op het Werk* - Federal Public Service Employment, Labour and Social Dialogue). A first example is a publication (188 pages - 2005) with tips for using machines and tools, where quite some attention is devoted to maintenance issues. The publication is part of the prevention culture the government wants to create. A second example is the project "SafeStart" aimed at a specific group, *i.e.*, young people starting in (often student) jobs. The project addresses this particular group with brochures and movies adapted to its interests. Safety legislations exist in every country and stipulate the basic legal requirements for workplace safety and maintenance. These legal requirements, however, do vary from country to country.

22.6 Safety Measurement

The term performance can be defined as the way in which someone or something functions and thereby accomplishes its purposeful objectives. In order to monitor and evaluate how well someone or something is doing ('performing'), performance needs to be quantified. The process of quantification of the performance can be broadly described as performance measurement (Neely). Performance measures

are important in business processes as they quantitatively let management know how well the business is doing, if goals are met, if stakeholders are satisfied, if processes are in control, if and where improvements are necessary. This objective also holds for safety performance measurement.

Safety performance measurement is important in industry as:

- It supports the monitoring and control of all safety related issues in the plant;
- It helps in the identification of areas that needs attention and improvement;
- It helps employees and management to focus their attention and resources to safety related aspects of the industry; and
- It helps in the control, management and improvement of the plant's safety.

To support measurement of safety in industry, a number of safety performance metrics, (commonly referred to as safety indicators), have been developed in both theory and practice. The number of safety indicators present in today's chemical process industry is overwhelming as discussed by Tixier *et al.* (2002). These indicators are categorized in several ways in literature, for example pro-active vs reactive indicators. Some of these classifications in the literature contradict each other. Some authors, like Kletz (1998) define pro-active as prior to the operational phase of an installation, while other authors like Rasmussen and Svendung (2000) define pro-active as prior to an accident. In this text, the definition of Rasmussen and Svendung (2000) is adopted, defining *pro-active* indicators as indicators before an accident and *reactive* indicators as indicators after an accident.

Another classification that is similar to reactive and proactive classification is the leading and lagging safety indicators (Van den Bergh and Butaye 2005). The lagging indicators are those that measure what has already happened, and in this case, with respect to safety violation or accidents. The lagging indicators thus provide the long-term trends of historical occurrences in the plant. They can therefore be referred to as reactive indicators. The lagging indicators are normally accurate in quantifying what happened in the past. For example, the analysis of the number of accidents can provide solution or conclusion for prevention of similar accidents in the future. They also have the comfort of seeing the safety trend already in motion. However, the information may come a little too late and with a heavy price to pay. For example, with the number of accidents as an indicator, the company has to wait until accidents happen to see where improvements are necessary. The lagging or reactive indicators thus have the disadvantage of not being able to identify and intervene safety hazards at an early stage.

The leading indicators are used to predict accidents before they happen. Moreover, they monitor the condition of the plant with regard to safety related issues in the plant. They also involve measurement of management efforts in preventing and mitigating accidents. These indicators can be referred to as proactive indicators. However, this category of measures has the disadvantage of not always being accurate. For example, leading measures like safety audit score, behavioural indicators or organization risk factors are highly dependent on people's perception of risks and accidents. Different auditors or safety inspectors can give varying scores for the same plant.

In this text, we classify safety performance indicators as lagging (reactive) and leading (proactive) indicators. Some examples of reactive safety indicators are accident rate, severity rate, lost time injury rate, accident cost, *etc.* The pro-active indicators are sub-divided into predictive/monitoring indicators and safety effort indicators. According to Korvers (2004), the monitoring indicators use actual events as a measure for the likelihood, while the predictive indicators predict the likelihood. However, there is no clear difference between the two and we therefore classify the two in the same category. Some examples in this category are safety deviations, near misses, accident free period, safety audit score, *etc.* The safety effort indicators try to quantify the management efforts directed towards safety improvements. The safety improvement efforts can be quantified in terms of safety audits, risk assessment, safety training, safety budget, *etc.* The intuition behind safety efforts is that they lead to improved safety and therefore less or no accidents. However, there is limited scientific research to prove the relationship between management efforts and plant safety results.

The classification of safety performance indicators with examples in each category is shown in Figure 22.5. Some examples of important safety indicators are given for each category. However, there are pros and cons associated with each indicator, though the details are not included in this text.

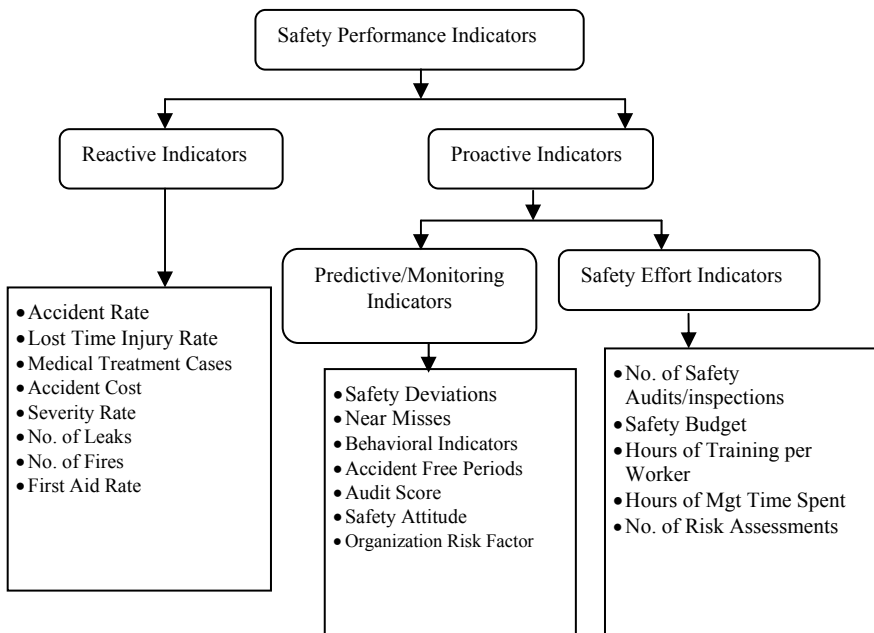


Figure 22.5. Examples and classification of safety performance indicators

References

- Anderson RT and Neri L (1990) *Reliability Centred Maintenance: Management and Engineering Methods*. Elsevier Applied Sciences, London
- API (1976) RP 520 Recommended Practice for the Design and Installation of Pressure-relieving systems in Refineries. American Petroleum Institute, New York.
- API (2007) *Safe Maintenance Practices in Refineries*. American Petroleum Institute, New York.
- Batson RG, Ray PS, Wan Q, Weems WH (1999). How preventive maintenance impacts plant safety. Dept. of Industrial Engineering, University of Alabama.
- Blanchard BS (1992) *Logistics Engineering and Management*. Prentice Hall, Englewood Cliffs, New Jersey.
- BSI (1984) BS3811 Glossary of maintenance terms in Terotechnology. British Standard Institution, London.
- Chemical Safety Board (2005) Isomerization Unit Explosion Report, BP Texas Oil Refinery Report. Fatal Accident Investigation Report, US. Approved for release by J. Mogford, Investigation team leader.
- Christensen JM and Howard JM (1981) Field experience in maintenance. *Human Detection and Diagnosis of System Failure*, Plenum Press, New York: pp 111–133.
- Craft ML (1991) Implementation of corporate contractor safety guideline. *Proceedings of the first International Conference on Health, Safety and Environment*, The Hague. Society of Petroleum Engineers: pp 457–464.
- Darby TF, Pickup L, Wilson J (2005) Safety Culture in Railway Maintenance. *Saf Sci* 43: 39–60.
- Dekker R (1996) Application of Maintenance Optimization Models: a Review and Analysis. *Reliab Eng and Syst Saf* 51: 229–240.
- Dept.of.Energy (1990) *The Public Inquiry into the Piper Alpha Disaster*. Chairman Hon. Lord Cullen, HMSO, London.
- Dhillon BS (2002) *Engineering and Safety*. World Scientific Publishing, New Jersey.
- Gero D (1993) *Aviation Disasters*. Published by Patrick Stephens, Sparkford, U.K.
- Goetsch DL (1996) *Occupational Safety and Healthy*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Goetsch DL (1999) *Occupational Safety and Health for Technologist, Engineers and Manager*. Prentice-hall, Englewood Cliffs, N.J Third Edition.
- Hale AH, Heming BHJ, Smit K, Rodenburg FG, Van Leeuwen ND (1998) Evaluating Safety in the Management of Maintenance Activities in the Process Industry. *Saf Sci* 28(1): 21–44.
- Hammer W (1976) *Occupational Safety Management and Engineering*. Prentice-hall, Englewood Cliffs, N.J Second edition.
- Heinrich HW, Peterson D and Ross N (1980) *Industrial Accidents Prevention*. McGraw-Hill, New York, 5th Edition.
- HSE (1987). *Deadly Maintenance*. Health and Safety Executive, HMSO, London.
- HSE (1999) *Reducing Error and Influencing Behaviour*. HMSO, Norwich, UK HSG48.
- http://en.wikipedia.org/wiki/Occupational_safety
- <http://www.osha.europa.eu>
- Hudson P (1999) Safety Culture – Theory and Practice. *The Human Factor in System Reliability – Is Human Performance Predictable?* Siena RTO MP-032, 1–12.
- Hurst NW, Bellamy LJ, Geyer TAW and Astley JA (1991) A classification system for pipework failures to include human and socio-technical errors and their contribution to pipework failure frequencies. *J of Hazard Mater* 26: 159–186.
- IDCON (2007) *Safety and Maintenance. Reliability and Maintenance Management Consultants*, NC, US.
- King R (1990) *Safety in the Process Industry*. Butterworth–Heinemann Publications, U.K.

- Kletz T (1992) *An Engineer's View of Human Error*. Institution of Chemical Engineers, UK.
- Kletz TA (1998) *Process Plants: A handbook for Inherently Safer Design*. Taylor and Francis, Washington D.C. 2nd ed.
- Koehorst LJB (1989) *Analyse van een selectie ongevallen in de Chemische Industrie uit de Databank FACT*. (Analysis of a selection of Accidents from the Database FACTS). SDU, Den Haag.
- Korvers PMW (2004) *Accident Precursors: Pro-active identification of safety risks in the chemical process industry*. Technische universiteit Eindhoven (phd Dissertation).
- Laing PM (1992) *Accident Prevention Manual for Business and Industry*. National Safety Council 10th edition.
- Latino CJ (1999) *Hidden Treasure: Eliminating chronic failures can put maintenance costs up to 60%*. Reliability Centre, report, Hopewell, Virginia.
- Levitt J (1997) *The Handbook of Maintenance Management*. Industrial Press, New York.
- Manson S (2003) *Improving Maintenance by Reducing Human Error*. Health, Safety & Engineering Consultants Limited (HSEC).
- Moubray J (1997) *Reliability-Centred Maintenance*. Second Edition. Butterworth-Heinemann, Oxford.
- Nakajima S (1989) *Implementing Total Productive Maintenance*. Ocat Business Books, St Catherines.
- Nowlan FS and Heap HF (1978) *Reliability Centered Maintenance*. United Airlines Publications, San Fransisco.
- NSC (1997) *Accidents Facts*. National safety Council, Chicago, US 1997 edition.
- NSC (1999) *Accident Facts*. National safety Council, Chicago, US.
- OSHA (1989) *Control of Hazardous Energy Sources*. Federal Register 54(169): 3664436687.
- OTA (1995) *Gauging Control Technology and Regulatory Impacts in Occupational Safety and Health: An Appraisal of OSHA's Analytic Approach*. OTA-ENV-635, September 1995.
- Parkes DJ, Jardine AKS (1970) *Operational Research in Maintenance*. University of Manchester Press, Manchester, UK.
- Pichot C (2006) *Sonte at securitie au travail: Les meties de la maintenance en premiere* (Health and Safety on the Work Floor: Maintenance Jobs). AFIM, Paris, France.
- Pinjala K (2007) *Dynamics of Maintenance strategy*. Katholieke Universiteit Leuven (Centrum Voor Industrial Beleid), Doctoral Dissertation.
- Pintelon L and Parodi A (2007) *Evolution of Maintenance Concepts*. In *Maintenance of complex systems: Blending Theory and practice*, Murphy, P. and Kobbacy, K. (eds), Springer-Verlang, Guilford.
- Pintelon L and Van Puyvelde F (2006). *Maintenance Decision Making*. Acco Publishers, Leuven, Belgium.
- Pintelon L, Gelders L Van Puyvelde F (1997) *Maintenance Management*. Acco Publishers, Leuven, Belgium. First Edition.
- Pintelon L, Gelders L, Van Puyvelde, F (2000) *Maintenance Management*. Acco Publishers, Leuven, Belgium. 2nd Edition.
- Raouf A (2007) *On Modelling Organizational Safety*. Proc Pakistan Acad Sci 44 (1): 73–81
- Rasmussen J and Svendung I (2000) *Pro-active Risk Management in a Dynamic Society*. Swedish Rescue Services Agency, Karlstad.
- Ray PS, Batson RG, Weems WH, Wan Q, Sorock GS, Matz S, and Cotnam J (2000) *Impact of Maintenance function on Plant Safety*. Safety Research. August 2000 Edition: pp 45–48.
- Roland H and Moriarty B (1983). *Safety System Engineering and Management*. Wiley & Sons, Inc, Canada.
- Russell PD (1994) *Management strategies for accident prevention*. Air Asia 6: 31–41.

- Safety Board (1984) Aircraft Accident Report: Eastern Air Lines Lockheed L-1011 in Miami, Florida. National Transport Safety Board, Washington D.C. Report No. 84-04.
- Stoneham D (1998) *The Maintenance Management and Technology Handbook*. Elsevier Science, Oxford, U.K.
- Takahashi Y and Takashi O. (1990) *TPM: Total Productive Maintenance*. Asian Productivity Organization, Tokyo.
- Tania M (2003) *Productive Safety Management*. Butterworth – Heinemann Publications, U.K.
- Tixier J, Dusserre G, Salvi O, Gaston D (2002) Review of 62 risk analysis methodologies of industrial plants. *J of Loss Prev in the Process Ind* 15: 291–303.
- Transport Ministry (1992) Report on the accident to BAC 1-11. Air Accident Investigation Branch, London, U.K: Report No. 1–92.
- US Army (1972) *Maintainability Guide for Design*. US Army Material Command, Department of the Army, Washington, DC(AMCP-706-134).
- US.Navy (1992) *Quality Assurance: Submarine Maintenance Engineering*. Joint Fleet Maintenance Manual, US Navy, Portsmouth, New Hampshire Vol. 5.
- Van den Bergh K and Butaye O (2005) *A Plant Safety Measurement Tool*. Katholieke Universiteit Leuven (Centrum Voor Industrial Beleid), Master Dissertation.
- Waeyenbergh G and Pintelon L (2002) A Framework for Industrial Maintenance Concept Development. *Int J of Prod Econ* 77: 299–313.

Maintenance Quality and Environmental Performance Improvement: An Integrated Approach

Abdul Raouf

23.1 Introduction

Womack *et al.* (1990) coined the term “lean production”. In the lean context, non-value-adding activity was viewed as any activity that does not lead directly to creating the product. The approach is based on reducing the non-value-adding activities which results in savings to the company. It has been reported that activities not adding value to the product comprise more than 90% of the total activity (Caulkin, 2002). Total Productive Maintenance (TPM) is an approach which aims at the total elimination of all losses, including breakdowns, equipment set ups, adjustment losses, minor stoppages, reduced speed, defects and rework and all major yield losses. It may be said that the ultimate goal of TPM are few equipment breakdowns and zero product defects resulting in ultimate utilization of production assets and plant capacity. Romm (1994) indicates that environmental benefits are involved in the lean implementation. A strong relationship between lean manufacturing and environmental improvement has been reported (Waldrop, 1999; Pojasek and Five, 1999; Florida, 1996; Hart, 1997). The foregoing suggests that maintenance quality, which essentially has a similar objective to lean manufacturing, has a strong relationship with environmental improvement.

Brah and Chong (2004) have reported that maintenance quality plays a major role in reducing costs and improving product quality. There are several indicators that measure maintenance performance (Duffuaa *et al.* 1999; Niebel, 1994). Some agree that ensuring the lowest possible risks of harming the environment is one of the major objectives of maintenance (Pickwell, 2001), yet this aspect is not covered by any of the maintenance-related performance measures currently in use.

This chapter proposes an integrated approach for monitoring maintenance quality and environmental performance simultaneously. First it briefly outlines the traditional approach for improving maintenance quality such as TPM, applications

of Deming's 14 points to maintenance quality, benchmarking, maintenance audit and using stakeholder's satisfaction level as a feed back. Second, it outlines the relationship between maintenance quality and environmental performance. An instrument for monitoring maintenance quality and environmental performance is presented.

23.2 Maintenance Quality

Maintenance quality is hard to define and a universally acceptable definition is lacking. The effects of poor maintenance quality are easily noticeable while the process of higher maintenance quality can go unnoticed. Maintenance quality is high when the production yield is at its peak without unplanned stops and the cost of maintenance is minimum. Stevens (2001) has suggested Repeat Job index as a measure of maintenance quality which is obtained by dividing the number of repeat jobs this year by the number of repeat jobs last year.

It is generally agreed that maintenance quality has a direct link to product quality. High maintenance quality may result in reducing down time of equipment. Properly maintained equipment retains its capability over a longer period of time and this results in scrap reduction. Repeat calls for repairing the same defect in a given machine are a clear indication of the maintenance quality. Each stakeholder normally develops its own measure of maintenance quality.

23.2.1 Improving Maintenance Quality

There exist many approaches for improving maintenance quality. This includes application of Total Productive Maintenance (TPM), use of Deming's 14 points to maintenance management, benchmarking and auditing. Maintenance systems have various stakeholders. The satisfaction level of these stakeholders can be used as an indication of the level of maintenance quality. Performance audits are also used to assess the current level of maintenance quality and assist continuous performance of maintenance quality.

23.2.1.1 Total Productive Maintenance (TPM)

Total productive maintenance (TPM) is a considered source of improvement for a company's performance and is a possible next step for adding to the benefits of total quality management (TQM) philosophy. TPM seeks to engage all levels and functions in a company to maximize the overall equipment effectiveness by involving workers in all departments and levels and functions from shop floor to senior executives. TPM is focused principally on keeping a plant more efficient, carrying out preventive, corrective and autonomous maintenance. TPM can increase the longevity of equipment, thus reducing the need to replace equipment. Complete details about TPM and its techniques are provided in Nakajima (1988).

23.2.1.2 Deming's 14 Points

Deming's method of improving quality which is essentially based on his 14 points is credited with higher quality products, higher volume of production and reduction in scrap and rework (Walton, 1995). In view of the similarities between TQM and TPM, Deming's points may as well be used for improving maintenance quality. The relevance of each of the 14 points to maintenance quality is shown in Table 23.1. An argument can be made for using these points as guidelines by the management to improve maintenance quality.

23.2.2 Benchmarking and Quality

Benchmarking may be defined as an external focus on internal activities, functions, or operations in order to achieve continuous improvement (McNair and Leiberfried, 1992). It may be considered as a systematic process for measuring "best practice" and comparing to company's performance in order to identify opportunities for improvement and superior performance (Bahrami, 1999).

The benchmarking process normally consists of the following steps:

- Selection of a comprehensive set of parameters for comparison;
- Selection of external sites for comparison based on performance;
- Comparison of own parameters with the world class measures; and
- Identification of areas of greatest improvements.

The comparison of own performance with world class measures leads to a prioritized array of optimizing changes directed to achieving best practice level of effectiveness. Relationship between benchmarking process and the way it fits into an overall process of continuous improvement is shown in Figure 23.1.

Benchmarks are the performance indicators which drive the continuous improvement process. Some of the Best Practice Benchmarks are shown in Table 23.2. Each category of benchmarks may vary from industry to industry and also with time. Maintenance quality may be improved by comparing the company's performance against the benchmarks. Teams are formed to identify significant deviations and the improvements to be incorporated. Spider charts are frequently used to compare the performance with benchmark objectives. To assess the current performance level of a company, audits are used as well.

Table 23.1. Deming's fourteen points application to maintenance quality

Deming's points	Relevance to maintenance quality
Create constancy of purpose towards improvement	There must be direct relation between resource allocation and high maintenance quality. When the desired level of maintenance quality is reached, resource allocation shall not be varied (reduced)
Adopt the new philosophy	Maintenance must have a top to bottom approach based on proactive measures rather than bottom to top based on reactive measures
Cease dependence on inspection	Craftsmen must be coached and trained, kept motivated and when a defect occurs, reasons for its occurrence must be identified
Move towards a single supplier for any one item.	Eliminate suppliers that cannot qualify with statistical evidence of quality. Life cycle cost must also be considered along with price
Improve constantly and forever	To better the quality and productivity and thus constantly reduce cost it is inevitable to reduce variations. Establishing and using Key performance indicators is essential to improve maintenance quality
Institute training on the job	If people are not trained adequately their performance will vary. To improve maintenance quality and productive quality, variance must be minimum. Methods to identify craftsmen who need training must be developed and training sessions regularly scheduled. Means of assessing effectiveness of training must be developed as well
Institute leadership	The aim of the leader should be to improve quality and productivity by ensuring that equipment, space needed, <i>etc.</i> , are available at the right time and that interruptions and unplanned work are minimized
Drive out fear	Management by fear is counter-productive in the long term, because it prevents workers from acting in the organization's best interests. Fear of losing job interferes with the ability to concentrate and gets in the way of satisfaction and pride that entails a job well done
Break down barriers between departments	Each department serves not the management, but the other departments that use its outputs. To arrive at best solutions to problems, knowledge from other departments should be included as the solution to a maintenance problem
Eliminate slogans, numerical goals and posters.	It's not people who make most mistakes – it's the process they are working within. Harassing the workforce without improving the processes is counter-productive. A well planned maintenance operation is a stable process and yields higher results
Eliminate management by objectives	Production targets encourage the delivery of poor-quality goods. Many aspects of maintenance work involve decision making and it is hard to measure time for decision making jobs
Remove barriers to pride of workmanship	Increasing number of maintenance jobs completed by a craftsman are likely to result in the increase in repeat jobs
Institute education and self-improvement	Since technology is changing rapidly, skills needed to maintain equipment must also change too. A predetermined number of total work hours have to be dedicated for training
The transformation is everyone's job	To improve maintenance quality, talents of each person involved with maintenance related work are needed. This minimizes maintenance work and maximizes ease of maintenance

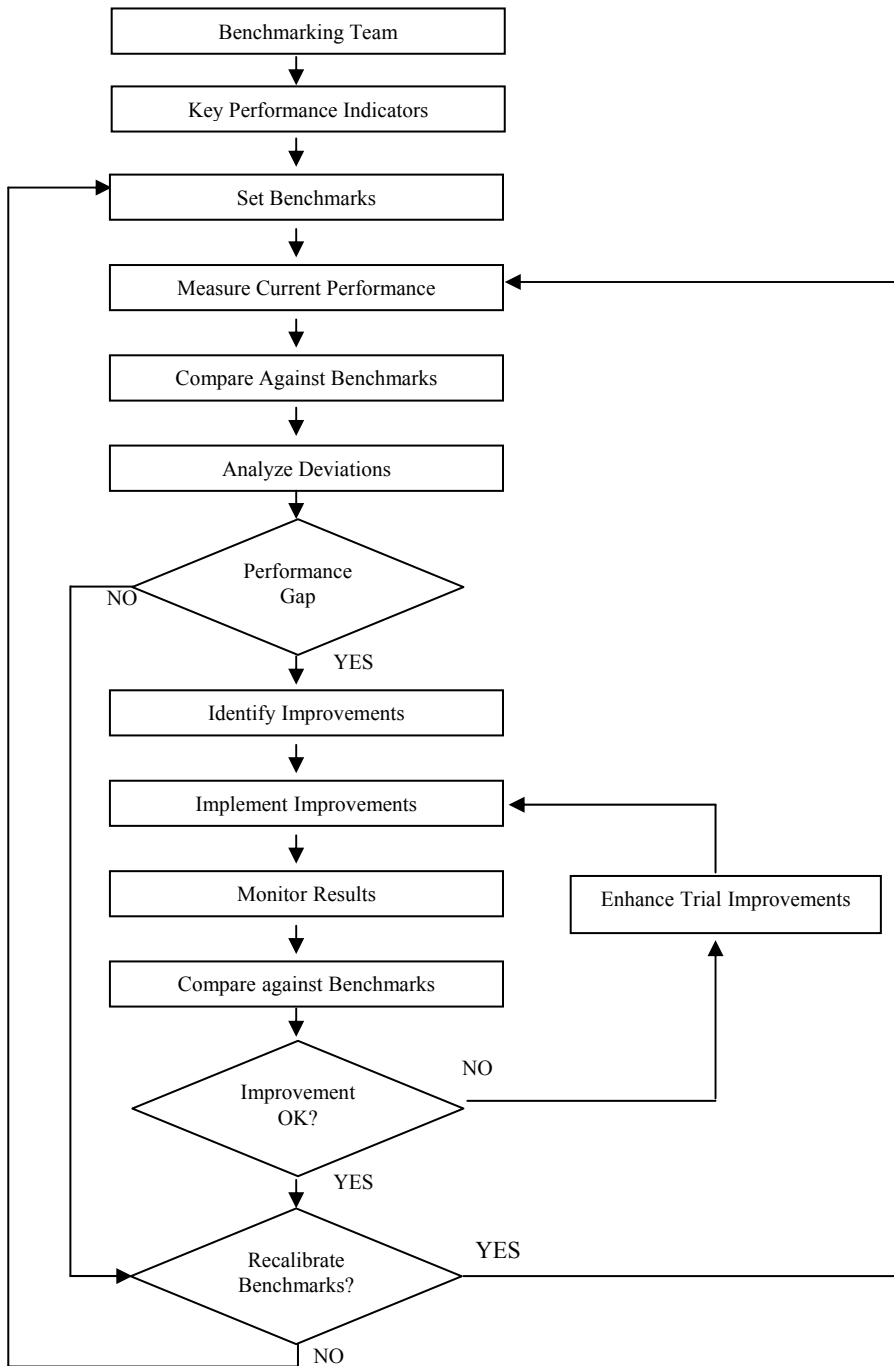
**Figure 23.1.** Benchmarking and continuous improvement

Table 23.2. Best practice benchmarking (after Bahrami, 1999)

BEST PRACTICE BENCHMARKS		
Category		Score
		Benchmarks
1.	Yearly Maintenance Cost	
	Total Maintenance Cost/Total Manufacturing Cost	< 10-15%
	Maintenance Cost/Replacement Asset Value of the Plant and Equipment	< 3%
2.	Hourly Maintenance Workers as a % of Total	15%
3.	Planned Maintenance	
	Planned Maintenance/Total Maintenance	> 90%
	Planned and Scheduled Maintenance as a % of Hours Worked	~ 85-95%
4.	Unplanned Down Time	~ 0%
5.	Reactive Maintenance	< 10%
6.	Run to Fall (Emergency + Non Emergency)	< 10%
7.	Maintenance Overtime	
	Maintenance Overtime/Total Company Overtime	< 5 %
8.	Monthly Maintenance Rework:	
	Work Order Reworked/Total Work Orders	~ 0%
9.	Inventory Turns:	
	Turns Ration of Spare Parts	> 2.8%
10.	Training:	
	For at least 90% of workers, h/year	> 40 h/year
	Spending on Worker Training (% of Payroll)	~ 4%
11.	Safety performance:	
	OSHA Injuries per 200,000 labor hours	< 2%
	Housekeeping	~ 96%
12.	Monthly Maintenance Strategies:	
	PM: Total Hours PM/Total Maintenance Hours Available	~ 20%
	PDM/CBM: Total Hours PDM/Total Maintenance Hours Available	~ 50%
	PRM (planned reactive): Total Hours PRM/Total Maintenance Hours Available	~ 20%
	REM (reactive, emergency): Total Hours REM/Total Maintenance Hours Available	~ 2%
	RNEM (non-emergency): Total Hours RNEM/Total Maintenance Hours Available	~ 8%
13.	Plant Availability: Available Time/Maximum Available Time	> 97%

23.2.3 Maintenance Audit

Maintenance system's performance can be improved by continuously monitoring it. The starting point in the design of any improvement program is the assessment of the current status of the system.

Duffuaa *et al.* (1999) has developed a two-step audit process. The first step is the scoring of essential factors in the maintenance system and the second step is to obtain an audit score.

Following are the factors that constitute the basis of the audit:

- Organization and staffing;
- Labor productivity;
- Management training;
- Planner training;
- Craft training;
- Motivation;
- Management and budget;
- Work order planning and scheduling;
- Facilities;
- Stores, materials and tool cabinet;
- Preventive maintenance and equipment history;
- Condition monitoring;
- Work measurement and incentives; and
- Information system.

The information and importance and impact of each factor on a maintenance system's productivity are explained in Duffuaa and Raouf (1996) and Duffuaa *et al.* (1999).

For carrying out scoring a set of questions for each factor is developed and the score is given based on the answers to these questions. To obtain an audit score (Maintenance Audit Index) the weight of each factor is determined. Combining these steps, the maintenance audit score can be obtained. Based on the results of this maintenance audit a continuous improvement plan for a maintenance system can be developed. The concepts and practice of audit, including operations audit, encompasses more than just increasing efficiency and effectiveness of plant. Poor maintenance quality can lead to increased pollution and environmental costs. Higher maintenance quality aids in minimizing such costs. Special attention has got to be paid to the equipment whose malfunction or poor maintenance can result in occurrence of events having significant environmental impacts. The maintenance planning and scheduling functions must include activities to prevent the occurrence of events which may cause significant environmental impact. Tracing the events leading to occurrence of well known disasters (like the Bhopal incident in 1984) reveals that minor events which could have been corrected during maintenance operations resulted in such major disasters (Raouf, 2007; Naryan, 2004).

23.2.4 Improving Maintenance Quality Based on Stakeholder Feedback

Maintenance activities are meant to provide services to various stakeholders within the company. Identifying such stakeholders and defining the service that each expects can help the company to receive feedback regarding the quality of service provided. Following are the typical stakeholders of a maintenance system:

- Operations/production;
- Purchasing;
- Maintenance management;
- Engineering;
- Top management;
- Accounting; and
- Storeroom.

Each stakeholder assesses maintenance quality as per its own performance measure. Table 23.3 shows stakeholders and the possible criteria for maintenance quality. A standard technique of measuring customer satisfaction may be applied. The feedback thus obtained from the stakeholders may be used to improve their satisfaction level. It will involve developing performance measures against each criterion for each stakeholder. For further details please refer to Hayes (1997). Needless to say, the higher the satisfaction level, the higher the maintenance quality. Continuous improvement framework for maintenance systems is shown in Figure 23.2.

23.3 Lean Manufacturing – Maintenance Quality Relationship

Poor maintenance quality has a negative impact on environmental performance. Poor maintenance quality is likely to result in over production, carrying extra inventory, extra use of transportation, defects produced, over processing, *etc.* Inadequate lean implementation also has a similar impact on environmental performance. This inter relationship and its effects are shown in Figure 23.3.

23.3.1 Basic Environmental Measure

Direct benefits associated with lean implementation and maintenance quality consist of material use, water use, energy use and waste generation. A list of suggested environmental performance matrices is shown in Table 23.4. This list is not exhaustive by any means.

To make significant improvement of environmental performance it is essential that the company's top management's commitment is visible and also that environmental performance be included in the training programs.

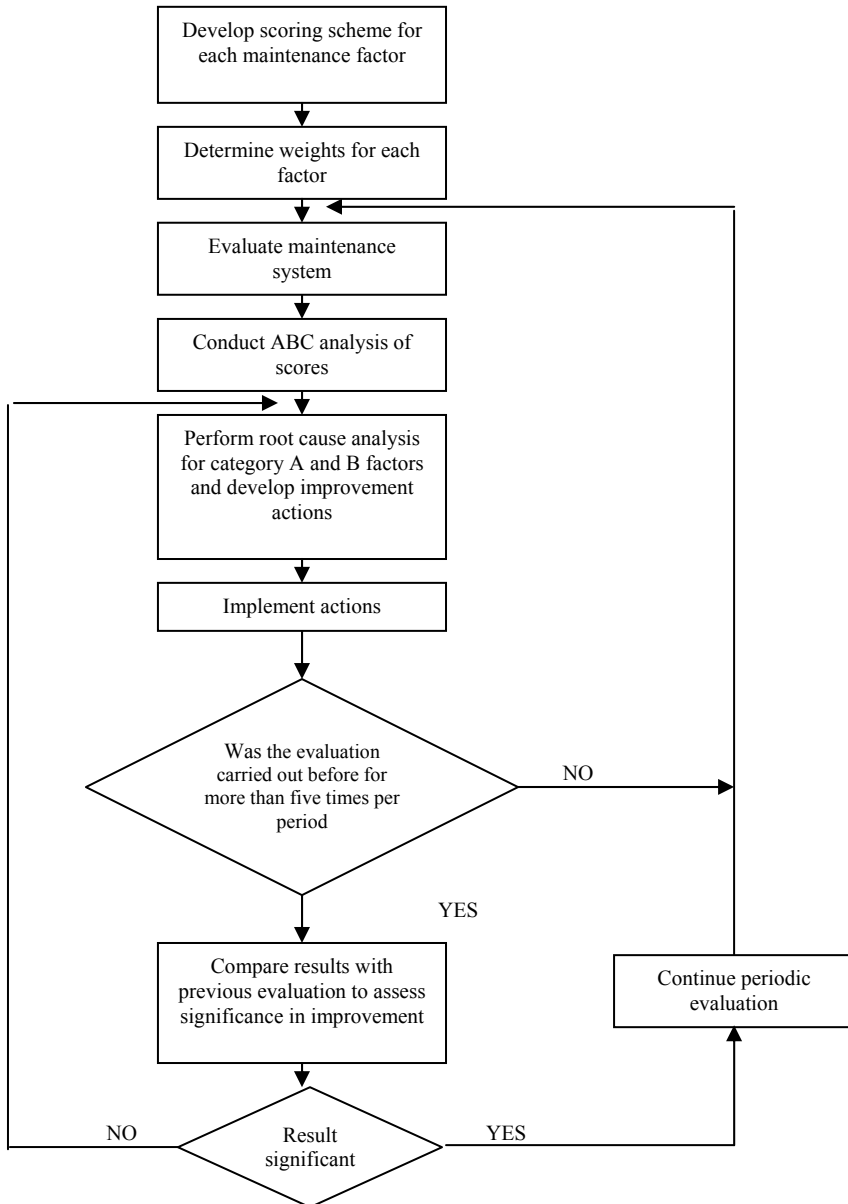


Figure 23.2. Continuous improvement plans for maintenance systems (after Duffuaa *et al.* 1999)

Table 23.3. Stakeholders and criteria for maintenance quality improvement

Stakeholders \ Criteria	Operations/Productions	Purchasing	Maintenance Manager	Engineering	Top Management	Accounting	Storeroom
Reduce the criticality and number of repairs	x		x				
Reduce downtime	x		x	x	x		
Increase equipment's useful life	x			x		x	
Increase operator, maintenance mechanic, and public safety	x				x		
Increase quality of output	x						
Reduce overtime		x				x	
Increase equipment availability	x						
Decrease potential exposure to liability					x	x	
Reduce number of standby units					x	x	
Increase control over spare parts and reduce inventory levels		x				x	x
Decrease unit part cost		x					
Lower overall maintenance costs through better use of labor and materials			x				
Lower cost/unit (cost per ton of steel, cost per cam shaft, cost per case of soda)	x				x		
Improve identification of problem areas to know where to focus attention	x			x			

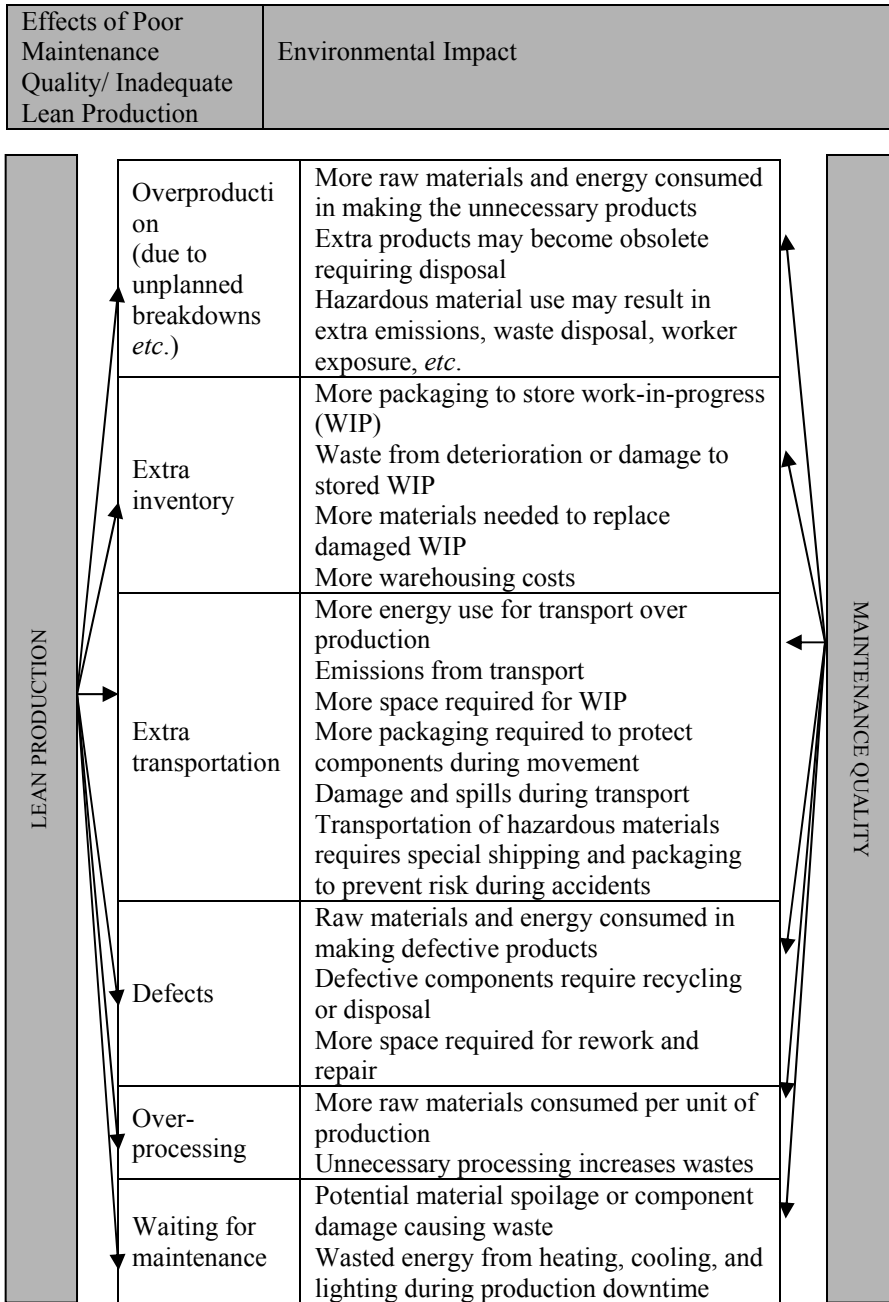


Figure 23.3. Effects of poor maintenance quality and its impact on environmental performance

Table 23.4. Suggested environmental performance matrix

Basic environmental measures		
Category	Definition	Unit of measure
Input measures		
Material use	Material used	Tons/year, pounds/unit of product, % materials utilization
Energy use	Any source providing usable power or consuming electricity transportation and non-transportation source	Specific to energy source such as BTUs or kilowatt hours, % reduction, energy use/unit of production
Water use	Incoming water from outside sources, <i>e.g.</i> , from municipal water supply or wells, for operations, facility use, and grounds maintenance	Gallons/year
Non-product output measures		
Air emissions	The release of air toxics	Pounds/year, tons/year
Water pollution	Quantity of pollutant in wastewater that is discharged to water source	Gallons or pounds/year
Solid Waste	Wastes (liquid or solid)	Gallons or pounds/year

23.4 Integrated Approach

Performance management is one of the basic requirements for determining the maintenance quality level and identifying the areas of improvement. Every company, more or less, develops its own maintenance performance indicators. A hierarchical approach to performance indicators has been suggested by Stevens (2001). On similar lines, an instrument to measure maintenance performance and environmental performance has been developed and is shown in Table 23.5. This may be considered as a guideline. Each company depending upon its type of operation may need a different set of measures. This instrument consists of areas and typical measures used. The areas are efficiency and effectiveness, tactical, functional and environmental. Targets can be set and a periodic review carried out to see the current status and identify areas where improvement is needed. This should be carried out by a team consisting of all the stakeholders if possible.

Previously maintenance manager’s problem was not having enough data to make an informal decision. With availability of computerized management systems *etc.*, the reverse is true. One feasible solution is to develop a knowledge base which should assist in providing actionable management information that is necessary for achieving the desired results.

A list of items where targeted figures are least met may be prepared in ascending order. This can assist in prioritizing the corrective action. While developing the priority list, due consideration of the criticality of the items covered must be kept in view.

Table 23.5. Typical performance indicators

Area	Typical indicators	Current assessment carried on target score of 5				
		100% (5)	75% (4)	50% (3)	30% (2)	< 30% (1)
Corporate Financial Efficiency and Effectiveness	• • • •					
Tactical	• • • •					
Functional	• • • •					
Environmental Performance	• • • •					
TOTAL SCORE =						

A list of typical performance indicators is as follows:

Corporate financial efficiency and effectiveness:

- Return on net assets; total cost to produce;
- Maintenance cost per unit produced;

- % of total direct maintenance cost that is break down related;
- % of WOs that are PMs;
- Maintenance related equipment downtime this year vs last year;
- Current maintenance costs vs those prior to predictive program;
- Number of repetitive failures vs total failures; and
- Overall equipment effectiveness combining availability, performance efficiency, and quality rate.

Tactical:

- % of total number of breakdowns that should have been prevented;
- Total of items filled on demand vs. total requested;
- Total Planned WOs vs total WOs received;
- PM hours performed by operators as % of total maintenance hours; and
- Number of equipment breakdowns per hour operated.

Functional:

- % of total WOs generated from PM inspections;
- % of total stock items inactive;
- % of total labor costs from WOs;
- % of total labor costs that are planned;
- % of total in plant equipment in CMMS;
- Training hours per employee;
- % of total hours worked by operators spent on equipment improvement;
- PdM hours % of total maintenance hours;
- % of failures where root cause analysis is performed;
- % of critical equipment covered by design studies;
- % of critical equipment where maintenance tasks are audited; and
- Savings from employee suggestions.

Environmental performance:

- Tons/years, lbs per unit of product, % of material utilized;
- BTUs or K.W., % reduction, energy use per unit;
- lbs/year, lbs per unit of product;
- Gallon/liter per year used;
- lbs/year; and
- Gallon/liter/year.

23.5 Conclusion

A new approach to maintenance quality has been developed. Various approaches for improving maintenance quality have been described. A linkage between maintenance quality and environmental performance alongwith measures of environmental performance have been presented.

Further work is needed to develop composite measures of maintenance quality and environmental performance.

References

- Bahrami K (1999) A maintenance improvement program benchmarking, Proceedings MARCONG.
- Brah SA and Chong WK (2004) Relationship between total production maintenance and performance. *Int J of Prod Res* 42: 2382–2401.
- Caulkin S (2002) Waste Not, Want Not. *The Observer*, September.
- Duffuaa SO and Raouf A (1996) Continuous Maintenance Productivity Improving Using Structural Audit. *Int J of Ind Eng Appl and Pract* 3: 151–160.
- Duffuaa SO, Raouf A and Campbell JD (1999) *Planning and Control of Maintenance Systems, Modeling and Analysis*. Wiley.
- Florida R (1996) Lean and Green, The Move to Environmentally Conscious Manufacturing. *Calif Manag Rev* 39(1): 80–105.
- Hart SL (1997) Beyond Green Strategies for a Sustainable World. *Harv Bus Rev* 75(1): 66–76.
- Hayes BE (1997) *Measuring Customer Satisfaction*. ASQL Quality Press.
- Makis V Jiang X, Jardine AKS and Cherry K (2000) A General Framework for Analyzing Maintenance Policies. *Maintenance Modeling and Optimization*, Ben-Daya, M, Duffuaa, SO and Raouf, A (Editors), Kluwer.
- McNair CJ and Leiberfried KHJ (1992) *Benchmarking; A tool for Continuous Improvement*. Omneo.
- Nakajima S (1988) *Introduction to Total Production Maintenance*. Productivity Press.
- Narayan V (2004) *Effective Maintenance Management; Risk and Reliability Strategies for Optimizing Performance*. Industrial Press.
- Nielbel BW (1994) *Engineering Maintenance Management*. Marcel Dekker.
- Pickwell JV (2001) *Maintenance Management Methodologies Excellence: Optimizing Equipment Life Cycle Decisions*, Editor Campbell JD and Jardine KS, Marcel and Decker.
- Pojasek R and Five S (1999) *A Tool That Prepares an Organization for Change*. Environmental Quality Management, Autumn.
- Raouf A (2007) On Modeling Organizational Safety. *Proceed of the Pak Acad of Sci* 44: 73–81.
- Romm J (1994) *Lean and Clean Management: How to Boost Profits and Productivity by Reducing Pollution*. Kodanasha International.
- Stevens B (2001) *Measurement in Maintenance Management in Maintenance Excellence: Optimizing Equipment Life Cycle Decisions*”, Editor Campbell JD and Jardine KS, Marcel and Decker.
- Waldrop G (1999) *Integrating the Elements for Sustainable Manufacturing*. Environmental Quality Management, Winter 1999.

- Walton M (1995) The Deming Management Method. *The Acad of Manag Rev* 13(1): 138–142.
- Womack J, Jones D, and Roos D (1990) *The Machine That Changes the World*. Simon and Schuster.

Industrial Asset Maintenance and Sustainability Performance: Economical, Environmental, and Societal Implications

Jayantha P. Liyanage, Fazleena Badurdeen, R.M. Chandima Ratnayake

24.1 Introduction

Sustainability performance appears to be one of the most influential concepts for managing modern businesses. Over the last few years it has drawn significant attention from many socio-political and socio-economical sources as the serious challenges encountered by both western and eastern societies were subjected to discussions and debates. This concept by far questions and challenges the fundamentals of commercial activities and its complex interactions with the environment external to an organization.

Sustainable development is defined as 'development that meets the needs of the present without compromising the ability of future generations to meet their own needs' (UNWECD, 1987).

To achieve sustainability, a commercial organization has to design and then adopt specific policies and procedures to guide and regulate its internal practices. These specifications and guidelines should help support or guide internal decisions and activities at various levels of an organization. One important aspect the organization needs to consider is the performance of the portfolio of assets, which in fact has a significant impact in this context. It implies that various processes at the production/manufacturing/process/infrastructure asset level have key roles and their performance levels are critical for sustainability compliance performance of an organization.

According to classical economic theories, asset maintenance is seen as a cost center. However, as managers have begun to realize the importance of intangibles and to re-examine industrial operations in terms of value added, asset maintenance is now seen not simply as a cost but rather as a process with significant potential to

add value. This newer view is bolstered by emerging concern for sustainability compliance. More recent publications that have brought this issue into open discussion include Liyanage (2003, 2007), Liyanage and Kumar (2003), Jawahir and Wanigaratne (2004), and Ratnayake and Liyanage (2007).

This chapter gives an overview of emerging sustainability issues and shows how the asset maintenance process plays an important role in sustainability compliance. It also elaborates on issues of quality and discusses best practices for guiding decisions.

24.2 Industrial Activities and Sustainability Trends

In theory, industrial economies exist to produce goods and services to improve the quality of life of their societies. However, this view is narrowly focused and does not consider the impact of these economies on other societies or on future generations. If the current consumption patterns in developed countries are extrapolated, it is estimated that the equivalent of three Earths (in terms of resources) will be required to provide the same quality of life for the rest of the world (Young, 2006). The situation will be further compounded with the projected increase in world population by another 2 billion by the year 2025, with most of the change taking place in developing countries (Holliday and Peppers, 2001). Given the impact of existing development patterns, based on intensive resource utilization with serious consequences for the ecosystem, there is naturally an increasing concern not simply with current problems but with the quality of life for future generations. This concern has brought about a more holistic and sustainability-oriented approach to growth and development.

In 1987, the United Nations World Commission on Environment and Development explicated the need to change course in economic growth and development through the Brundtland Commission's report. In the report titled '*Our Common Future*' the committee recommended the pursuit of 'sustainable development' (UNWECD, 1987), and emphasized the cautious use of natural resources with minimal impact on the ecosystem to meet the needs of all people; a move towards improving the triple bottom-line (TBL) namely (see also Elkington, 1998);

- Economic prosperity;
- Environmental protection; and
- Social equity.

Ever since, sustainable development has become a priority for more and more private and public sector organizations, guiding their policy decisions and strategic planning.

Today, many companies have begun to take initiatives to adopt a "*sustainable business policy*", or at least see the need to do so (Liyanage, 2003). The challenge to businesses is to keep operating profitably yet sustainably. On the other hand, the challenges to governments are to legislate and regulate commerce and consumption. However, despite the worries of businesses and the lack of political courage by governments, the stress on natural systems is no longer something that

can be ignored (WBCSD, 2004). Globalization has made economies more interdependent and interconnected, increasing cross border trading and transportation. The mobility of people, goods and other resources around the world has increased tremendously (Doering *et al.* 2002). To be competitive in the global market, companies and governments are now forced to seek global standardization of ecological and social standards (Keijzers, 2002) and they are globally answerable on the quality of their products and operations. Therefore, businesses have to be more transparent in their activities and take substantive responsibilities for sustainable ecological, economic and social development (Keijzers, 2002).

Consumers, too, are taking a more active role in insisting on corporate compliance, for instance to health and safety standards, eco-conservation and, in general, to ensure good corporate citizenship (Keijzers, 2002). Therefore, human and societal values now have a more significant impact on corporate policy and decision making. In addition, more and more non-governmental organizations (NGO) are acting to make business and governmental activities more transparent and more accountable to legislation in the interest of all stakeholders (WBCSD, 2006). It is likely that societies in many developed countries will soon begin to view sustainability as a new form of value (Elkington, 1998) and companies will be forced to respond to growing stakeholder insistence by formulating a coherent and fundamental strategy for sustainability rather than doing some *greenwashing* (CorpWatch, 2008). It appears that stakeholders must be engaged in business decision making to ensure expectations and motivations of all are congruent and aimed at achieving sustainability (Elliott, 2005).

Although sustainability has become 'a board-level agenda item' as noted by Elkington (1998), in many leading companies, achieving sustainable development will not be easy as it requires a major shift in the paradigm for strategic planning and operations management. Traditional approaches, based on 'doing the same things but better,' must be replaced by innovative approaches that do things differently, *i.e.*, use fewer resources to meet human needs while ensuring social equity and environmental protection (Dormann and Holliday, 2002). Pursuing growth strategies with piecemeal increments of sustainable activities added-in makes it even difficult for companies to achieve and maintain competitive advantage while being a good corporate citizen from a sustainability point of view. In the manufacturing sector, applying the 6R concept of *reduce, reuse, recycle, recover, redesign and remanufacture* can enable improving on the lean (waste reduction-based) and green (environmentally benign) manufacturing strategies to more holistic, *sustainable manufacturing* to achieve exponential growth in stakeholder value (Jawahir and Dillon, 2007) as shown in Figure 24.1. This innovation-based development will benefit all stakeholders and enable companies to depend more on human ingenuity, than exploitation of resources (Holliday and Peppers, 2001), to achieve the TBL.

Companies now appear to acknowledge the importance of pursuing sustainable business policies to ensure corporate success. Therefore, it is timely to begin incorporating sustainability thinking into all business operations, including asset maintenance which is a significant but relatively less known contributor to enhance enterprise sustainability. Sustainable asset maintenance focuses on prolonging the useful life of the assets and ensuring higher productivity from the systems they are

part of, through methods that achieve economic, eco-friendly and socially equitable goals. Thus, within the sustainability framework, sustainable asset maintenance has emerged as a vital ingredient to attain the status of a sustainable enterprise.

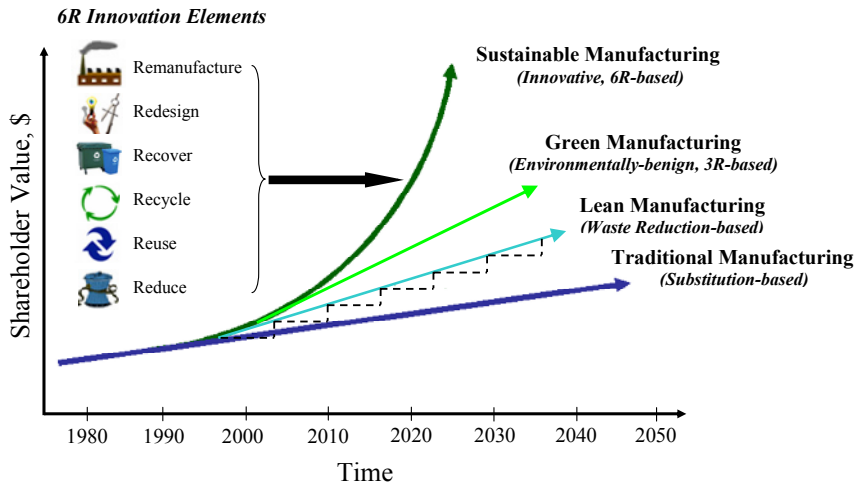


Figure 24.1. Sustainable manufacturing for the twenty first century (adapted from Jawahir and Dillon, 2007)

24.3 Sustainability Performance in Perspective

There is much agreement among researchers and practitioners on the importance of integrated *performance* of industrial assets for competitiveness in production, manufacturing and service organizations. However, the understanding of performance evaluation aspects of industrial assets with respect to business strategy, in particular with due focus on sustainability performance (SP), remains relatively undeveloped. To date, little data is available to permit assessment of:

- How extensively the use of performance measurement techniques complies with SP requirements at the corporate strategy formulation level;
- How these techniques have spread through plant/facility levels of organizations;
- What factors have influenced their diffusion; and
- How these techniques affect the overall organizational performance.

The present general agreement on the need to measure asset performance towards SP has not yet led to the development of a systematic process for determining appropriate measurements (indicators).

As shown by Ghalayini and Noble (1996), research on performance evolved through two phases. The first was the cost accounting orientation that was strongly criticized for encouraging short-term thinking (Banks and Wheelwright, 1979;

Hayes and Garvin, 1982; Kaplan, 1983) and for its failure to measure and integrate all the factors critical to business success (Kaplan, 1983, 1984). The second phase was associated with the growth of global business activities and the changes brought about by such growth. It called for development of better-integrated performance management systems stressing the importance of non-financial measures (Johnson and Kaplan, 1987, McNair and Mosconi, 1987, Santori and Anderson, 1987). Subsequently, some frameworks, which attempted to present a broader view of performance measurement started to appear (Cross and Lynch, 1988–1989, Khadem; 1988).

Despite financial issues remaining a significant factor, a wave of environmental concerns was experienced worldwide after the late 1960s (Caldwell, 1989). This gained further attention during the 1980s and 1990s becoming an independent subject of study. The mode of response to environmental demands was mainly reactive in the 1960s and early 1970s, with a substantial response from industry. Corporate pollution prevention plans and environmental management systems were introduced during this era. Then during the 1990s, a more comprehensive approach for assessing environmental costs based on full-cost accounting (Committee on Industrial Environmental Performance Metrics, 1999) and pollution prevention means for reducing pollutants (Oldenburg and Geiser, 1997) emerged. The underlying philosophy was to deliver the same output with a lesser environmental burden introducing methods referred to as ‘clean technologies’.

During the 1970s many organizations also began developing standards for corporate social accounting (Epstein, 1996). Despite the fading interest on the subject matter in the 1980’s, the efforts to measure and report social performance have resurfaced in the last few years (see, for instance, Daniel, 2005). This change has been in response to the need for societal indicators to evaluate sustainability and also to better communicate the impact of business operations. As a result, the Council on Economic Priorities (CEP) proposed SA 8000, a social accountability standard designed to follow in the path of other “quality” standards. CEP hopes that, similar to ISO 9000 and ISO 14000, SA 8000 will become the *de facto* standard for evaluating the quality of a company’s social performance. However, even if SA 8000 makes significant advances in standardizing the evaluation of corporate commitment to human issues, such as worker safety and equality, it covers only a limited subset of those required to ensure sustainability (Ranganathan, 1998).

Subsequent to the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in 1992 (or the “Earth Summit”, which called for a shift from “talk” to “action”), there has been an increased emphasis on the simultaneous consideration of economic growth, environmental protection, and social equity in business planning and decision-making (Schmidheiny, 1992). Ten years after the Rio Conference, in 2002, the World Summit on Sustainable Development (WSSD) was held in Johannesburg, where the business community also pursued to take an active role by launching a Business Action for Sustainable Development (Holliday *et al.* 2002).

An organization without a systematic way of understanding what it has or has not achieved is unlikely to succeed, irrespective of its aims or determination (Zadek *et al.* 1997). Thus, it is vital to set targets and measure performance

towards them for continuous improvement and development; the process must also generate information to identify gaps in performance for future action (Warhurst, 2002). Further, target definition and performance measurement needs to be pursued along all three sustainability focus areas. However, while financial performance measures are well established and serve as a basis for decision-making, other aspects—environmental and social—remain to be more informal.

Thus very few, if any, companies can respond to the question:

“Which of our products, processes, services, and facilities are in compliance with sustainability needs?”

Answering this question requires the ability to model and assess sustainability in an adequate manner, for which widely accepted or mandated standards are lacking. Sustainability performance measurement criteria are not similar to the conventional measures used for business performance assessment. This is because sustainability is a complex and multi-faceted concept, covering a broad variety of topics from habitat conservation, to energy consumption, to stakeholder satisfaction and financial results. Further, sustainability performance measurement requires extending beyond the boundaries of a single company and need to address the performance of both upstream suppliers and downstream customers in the value chain.

At the turn of the new millennium, many leading companies in the U.S., Europe and Japan began responding to the challenges of global population growth and environmental pressures by adopting a commitment to “sustainability” (Hart, 1997). Various terms such as sustainable development, sustainable growth, sustainable products, sustainable processes, and sustainable technologies have been used to describe this area of interest that has drawn the attention of business leaders. Several have launched proactive programs that include life cycle accounting, design for eco-efficiency, community outreach, clean technology development, and a variety of other initiatives. The motivations for underlying efforts are not purely altruistic, because recent research has demonstrated that pursuit of sustainability cannot only result in environmental improvements and societal benefits, but can also increase economic value for the firm (Kiernan and Martin, 1998; Dixon, 1999).

By 1992 there were already more than 70 definitions for sustainable development (Holmberg and Sandbrook, 1992). The International Institute for Sustainable Development (IISD) has subsequently suggested that businesses can gain a competitive edge, increase their market share, and boost shareholder value by adopting and implementing sustainable practices. This can be done by companies (see also Deloitte and Touche, 1992; Liyanage, 2003):

“Adopting business strategies and activities that meet the needs of the enterprise and its stakeholders today, while protecting, sustaining and enhancing the human and natural resources that will be needed in the future”.

Despite the much discussed publication by Elkington (1998) on the TBL, relatively little work has been done to explore how this concept impacts

performance at the plant/operational level of a business. The TBL framework presented complex aspects related to sustainability performance in a much simpler industrial context, and drew attention to how corporations manage and balance their responsibilities towards achieving a sustainable business. Some of these issues have also been discussed by Van Dieren (1995).

Most organizations have in fact begun to focus mostly on macro-environmental features of commercial activities. However, stakeholders are now demanding proof of the “sustainability” performance of operational initiatives. The sustainability consideration of such operational initiatives in industry need more focused methods that can measure both the beneficial and adverse impacts associated with plants, facilities or operations, *i.e.*, to assess to what extent the plant level operational initiatives, facilities or operations are aligned with the principles of corporate sustainability strategy. Such an approach, in the first place must help business process to integrate a company’s strategic planning into day-to-day operations.

24.4 Sustainability Performance Framework: From Business to Asset

As Daly (1990) points out, sustainable development is an inherently vague but a compelling concept, which remains difficult to express in concrete, operational terms (Briassoulis, 2001; Liyanage, 2003). On the other hand, putting this concept into practice has enlightened some of the best minds in leading global corporations.

As aforementioned, the major focus of sustainable performance is integrating the performance; in economic, social and environmental terms simultaneously. Hence the domain of performance can be conceptualized as being concerned with:

- Economic performance that strive to achieve economic objectives;
- Social performance that strive to fulfill of social objectives; and
- Environmental performance that cater to realize environmental objectives.

Through the World Summit 2002, it was revealed that all three pillars of the tripartite world would have to work together in partnerships to solve the challenges and to achieve true sustainability (Holliday *et al.* 2002). Such insights have contributed much to bring stakeholders within the sustainability concepts and subsequently to underline their impact on an organization’s competitive performance. It implies that the gradual organizational transition to adapt sustainability practices needs to be managed with due focus on and commitment to stakeholder requirements. This in fact poses a challenge to many organizations (Figure 24.2).

In order to overcome sustainability compliance challenges, organizations have to pay attention to and resolve some strategic questions as:

- How to address environmental and societal aspects of the organization while securing the economic performance?
- How to model, mandate, and govern complex business processes, preserving the integrity of value chain? and

- How to cascade business policies to asset operations to have positive business impact?

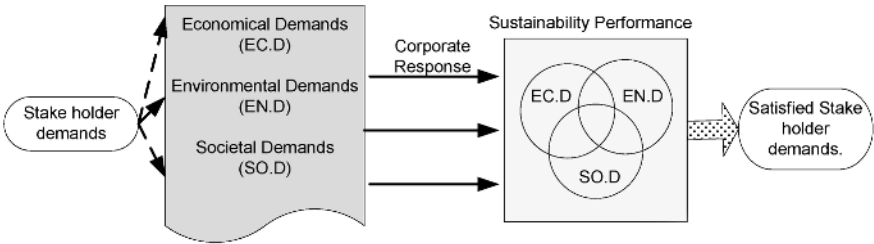


Figure 24.2. The sustainability performance challenge at present (see also Ratnayake and Liyanage, 2007)

In fact, within existing sustainability performance frameworks of commercial organizations, conscious and sensible monitoring and verification of systems’ and processes’ performance is still a complex issue. This is due to the ill-defined nature of the concept as mentioned above. Besides, a range of sensitive information needs to be integrated into managerial frameworks, which can be a daunting and a resource-consuming task. In order to make decisions and to allow sensible communication, the inherent complexity needs to be addressed through smaller operational units where advanced data management and solutions provide a holistic information and knowledge management platform.

Ideally, for incorporating business sustainability in operational terms, following three distinct levels within an organization can be subjected to change:

- The strategic level;
- The process or methodology level; and
- The Operational level.

This is also shown in Figure 24.3.

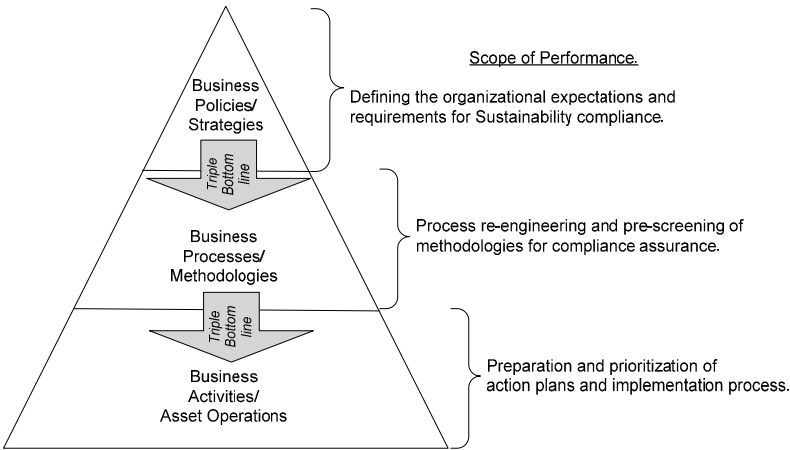


Figure 24.3. Introducing sustainability concepts to the asset portfolio of an organization

Despite the fact that corporate sustainability issues are still under discussion and development, the question, '*how does one distinguish a "sustainable" asset from one that is not*' poses new challenges for SP and SPM methodologies for industrial sectors, challenging the traditional scope of asset performance and performance management. Some of the difficulties that arise include:

- Lack of consensus on a pragmatic definition of sustainability at the plant level;
- Breadth of scope of sustainability issues, many of which are beyond the firm's control;
- Potentially large amount of information required for evaluating plant/process or operations sustainability; and
- Difficulty in quantifying the complex aspects of sustainability at the plant level.

The challenge of aligning operational processes with the corporate sustainability policies would be one of the most known difficulties. However, trying to achieve this type of alignment raises number of interesting non-conventional issues at the corporate and plant/operational level.

Corporate level issues include:

- Establishment of appropriate company policies and incentives;
- Modification of existing business model and policies;
- Decisions and procedures on capture and dissemination of sustainability knowledge via training and information management; and
- Achievement of consistent practices across diverse business units.

Plant or facilities related issues, on the other hand include for instance:

- Implementation of various engineering strategies, *e.g.*, modifying the material composition of products so that they generate less pollution and waste; or
- Changing the assembly requirements so that fewer material and energy resources are consumed per product unit; as well as
- Systematic adoption of sustainable design guidelines, metrics, and tools, *etc.*

It appears the pressure is gradually mounting on corporations to align all activities and operational processes with the principles of sustainable development (Keeble *et al.* 2003). This, to a large extent, relates to the deterioration of global life-support systems and the modernization of social systems imposing time limits for proactive actions. Consequently incorporating business sustainability in operational practices would be one of major effort to mitigate the former challenge. When it comes to SP of industrial assets, following three distinct levels within an organization can be subjected to change:

- Plant manufacturing/production strategy;
- Processes/methodologies; and
- Operations and activities.

Primarily, under the plant strategy one has to pay attention to distinguishing stakeholder satisfaction (*i.e.*, “*who are the key stakeholders and what do they want and need?*”) and strategies (*i.e.*, “*what strategies have to be put in place to satisfy the wants and needs of the stakeholders?*”). Based on the former, the next layer focuses on how to make the asset processes be aware of sustainable practices and what type of methodologies need to be adopted to help-support the processes to achieve sustainability performance. Under the operations and activities, on the other hand, one has to seek ways for bridging the gap, *i.e.*, what capabilities and resources are needed to operate and enhance the processes. Figure 24.4, elaborates how it is cascaded at each level while acting in accordance with institutional (stakeholder and projections on plant performance) and organizational characteristics (roles of parent company and demands for the plant).

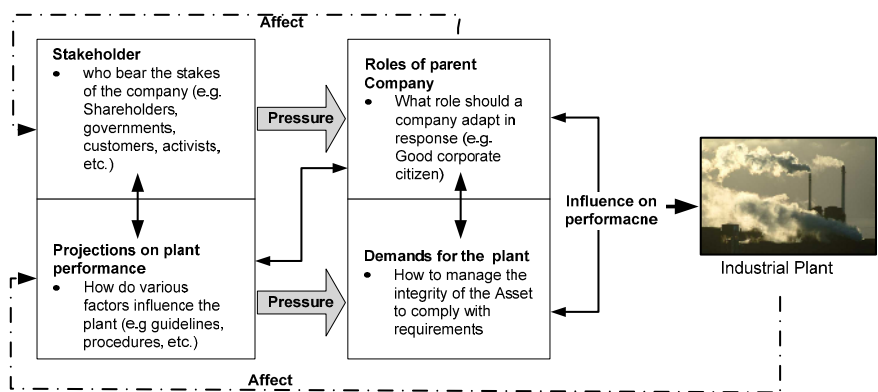


Figure 24.4. A model of institutional pressures moderated by parent company requirements to govern plant performance characteristics (also see Ratnayake and Liyanage, 2007)

The framework in Figure 24.4 in fact illustrates the drivers of sustainability of industrial activities. The principal issues (*stakeholder, roles of parent company, projections on plant performance, and demands for the plant*) provide the framework for the industrial sectors to be conscious about their own sustainability actions in response to stakeholders, thorough better plant/facility management practices.

The subsequent demands that are imposed on the plant can briefly be illustrated as in Figure 24.5.

Certainly, progressive and successful organizations have to adopt the practice of sustainability, and have to initiate strategy, tactics (processes/methodologies) and operations to remain competitive in the modern era. Nevertheless, for many businesses, this emerging perspective has been clouded by uncertainty over the linkage between perceived ‘external’ societal and institutional initiatives and ‘internal’ management strategies. There is a question about whether these processes can operate parallel to or in conjunction with each other. Changing

sources of corporate value indicate that these processes should be inclusive of each other.

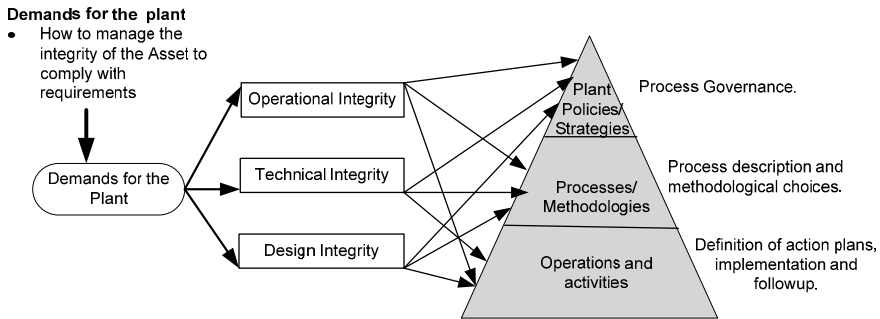


Figure 24.5. Measurement of plant performance characteristics (also see Ratnayake & Liyanage, 2007)

Focusing on SP, the corporate strategy can identify the plant strategy and its performance capabilities, in terms of, for instance:

- Supply chain and Logistics (service level and lead time);
- Economic value-added models (efficiency, costs and working capital);
- Human resources (including personal competence);
- Quality concepts (product revision, scrap and customer claims);
- Compliance verification programs (cost, environmental and societal effects); and
- Behavior and attitude enhancement strategies (*e.g.*, empowerment for self-control and self policing in peoples' management of natural resources, cultural identity, commonly accepted standards for honesty, laws, disciplines,..., *etc.*).

These capabilities represent some of the ideal tasks that should be incorporated to support the corporate strategy. To keep the elements intact, this calls for an integration strategy that should enable the holistic realization of competing objectives through proper monitoring and coordination of strategic, tactical and operational/activities. Such an integrated strategy can also lay the bedrock to provide a structured framework to manage information and knowledge to retain consistency in performance.

In fact the setback that most companies are facing today is the lack of such elaborative frameworks that allow development of assets for sustainability compliance. This implies that there is an inherent need for comprehensive SPM framework for industrial assets to address, check and balance these economic, environmental and social aspects at plant level. Although there has been proliferation of management systems, accounting, auditing and reporting standards, they were mostly focusing on selected issues without promoting a holistic approach. The need to integrate the TBL issues at each level, although realized by

many, leave them with a question of what framework is needed and really how to do and go about it to reduce the commercial risks and to enhance value-added (see discussions in Liyanage, 2003).

24.5 Defining Maintenance Custodianship Within an Asset's Sustainability Performance

Every individual asset comprises a number of technical processes that are key to operating the plant/facility in accordance with both commercial and legislative requirements. These core asset processes can vary depending on the product and/or operational characteristics of a given industrial asset. For instance, for a manufacturing facility the major processes can include:

- Product development;
- Production operations;
- Product and process quality control;
- Plant maintenance and process modifications; and
- Logistics and inventory management, *etc.*

These may vary somewhat in other industrial settings. For example, an oil and gas production complex has the following major processes:

- Drilling and well operations;
- Reservoir management and production;
- Operations and maintenance;
- Logistics and asset support services; and
- Asset development and modifications.

In a generic sense, every technical asset process has a specific role in the asset's sustainability compliance performance. These roles depend largely on the function that the technical processes have within the asset's operational framework (Figure 24.6).

This entails that the process level impact on an asset's sustainability performance can take different forms and magnitudes. For example, the product development process is concerned with the functional, aesthetic, and other characteristics of a specific product that goes to a particular market segment, while plant maintenance and process modifications is concerned with the technical excellence and safety integrity of systems and equipment that are critical for uninterrupted daily production operations.

For a very long time, plant/facility maintenance process has been regarded as a principal cost element within an asset performance framework. This postulates that the asset and business level contribution of maintenance process has often been subjected to a review solely based on a short-term financial impact. This is mostly seen expressed as a percentage of operating expenses (OPEX). However, over the last few years some attention has been drawn to exploring the role of maintenance process on a much wider scale. Some early work in this context has made efforts to

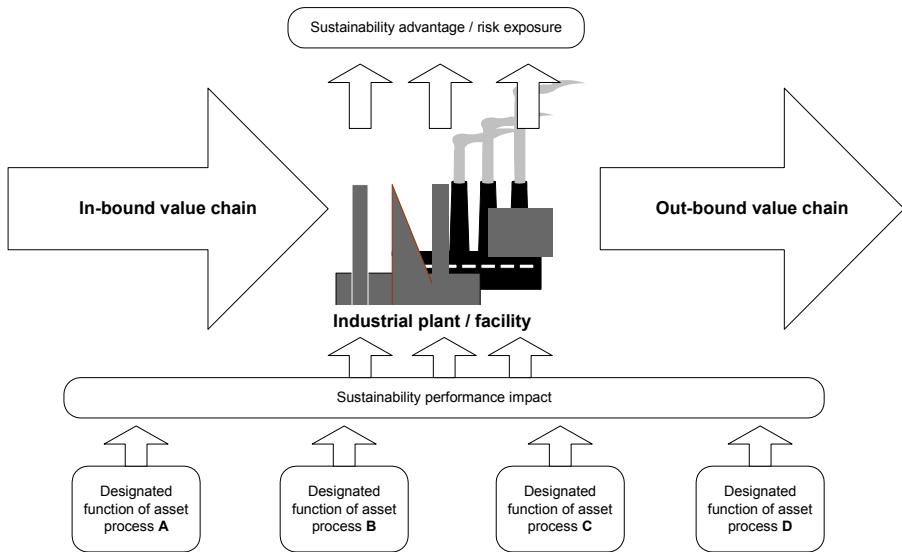


Figure 24.6. Asset processes have specific role-plays in the sustainability performance of a given industrial plant/facility

communicate the contribution of maintenance to profitability of production/manufacturing operations. The common denominator in such exercises has mostly been the maintenance impact on the systems' or equipment uptime (or alternatively on the production availability). Quantifiability of losses in financial terms in the event of production unavailability (or equipment downtime), for instance due to poor maintenance procedures, unattended maintenance work orders or work order backlog, *etc.*, have contributed much to these developments.

In essence, maintenance can be regarded as the discipline that directly affects and thus is accountable for the technical condition of an asset. This is so, regardless whether it refers to a production, manufacturing, or process asset, or an infrastructure asset. Thus, specification of a technical condition (and also in fact the safety integrity) of an industrial plant/facility is the principle basis for defining the *custodianship* of the maintenance process (Figure 24.7). The term '*custodianship*' here implies *the inherent responsibility designated to maintenance process for ensuring an acceptable condition of physical equipment, systems, and industrial facilities.*

The technical condition of a plant/facility can formally be expressed in terms of:

- Plant availability (on demand);
- Systems and equipment reliability; and
- Overall equipment effectiveness (OEE).

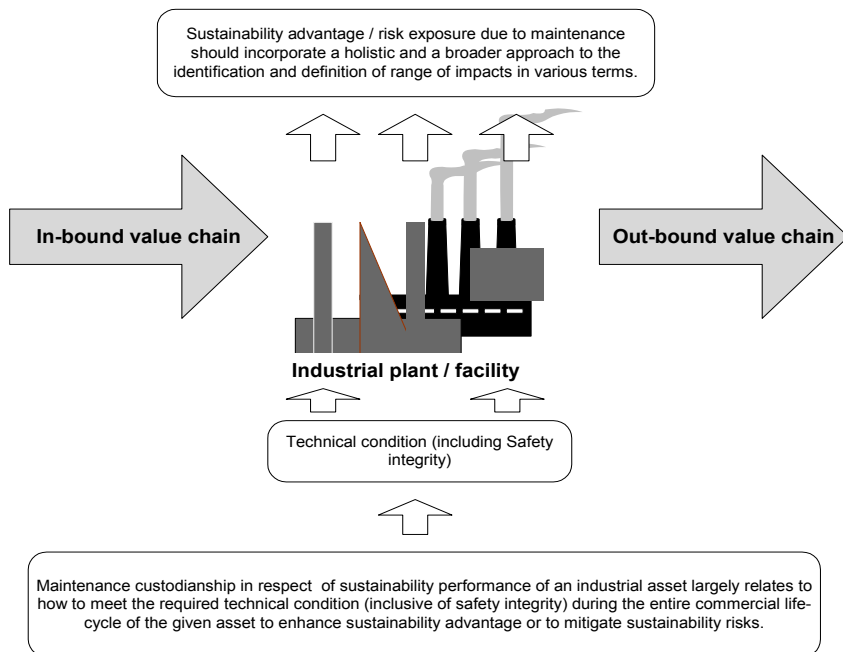


Figure 24.7. Framework for identifying and defining asset maintenance *custodianship* with respect to an asset's sustainability performance

Formally, there are two specific issues related to maintenance process that need to be addressed in sustainability performance context, *i.e.*;

- What is the nature and level of performance impact of the asset maintenance process during the entire commercial life-cycle of a given industrial asset.
- What is the nature of unwanted consequences that are likely to take place if the required/specified technical condition cannot be met through an effective and an efficient maintenance practice. (And also, what are the range of benefits that can be claimed if the required/specified technical condition can be met through an effective and an efficient maintenance practice).

More conventionally, maintenance *custodianship* is often seen narrowly defined in terms of systems or equipment faults/failures under given conditions during the operational phase of a facility. This conventional view arises on the basis of the fact that, unless otherwise it may expose the plant to operational risks only in terms of:

- Production anomalies (*e.g.*, quality shortfalls and/or reduced production); and
- Operational budget overruns (*e.g.*, due to hidden breakdowns, excessive resource consumption, idle times, *etc.*)

However in the wake of sustainability issues various other aspects take precedence in respect of overall asset performance. One can pay specific attention to the holistic maintenance impact on an asset's sustainability performance, through a relatively more thorough and broader analysis of underlying issues, as highlighted in Table 24.1.

Such an impact assessment, either in terms of *gains* or *losses*, can be made with reference to a number of key issues that are important for excellent operational performance of a plant/facility. Some examples are illustrated in Table 24.2.

Furthermore, the term 'overall asset performance' implies that there is an inherent need to focus on the maintenance (or maintainability) impact during the entire commercial life of an asset (so-called 'life-cycle' performance). Such an impact can be assessed in a number of terms, apart from being purely financial. This requires due analysis of loss potential or risk exposure if adopted maintenance programs and activities fail to meet the required technical condition.

Commercial life of a given industrial asset, in a more generic sense, can be divided into three major stages, namely:

- EPCIC (engineering, procurement, construction, installation and commissioning);
- Operational; and
- Decommissioning (or divestment).

Table 24.1. The basics for assessing maintenance impact on an asset's sustainability performance

Assessing impact in terms of <i>gains</i>	Assessing impact in terms of <i>losses</i>
What is the level of <i>financial impact</i> arising from excellent technical condition of systems/equipment of an asset due to effective and efficient maintenance practices?	What is the level of <i>financial impact</i> arising from poor technical condition of systems/equipment of an asset due to ill-defined and/or poor maintenance practices?
What is the level of <i>social impact</i> arising from excellent technical condition of systems/equipment of an asset due to effective and efficient maintenance practices?	What is the level of <i>social impact</i> arising from poor technical condition of systems/equipment of an asset due to ill-defined and/or poor maintenance practices?
What is the level of <i>environmental impact</i> arising from excellent technical condition of systems/equipment of an asset due to effective and efficient maintenance practices?	What is the level of <i>environmental impact</i> arising from poor technical condition of systems/equipment of an asset due to ill-defined and/or poor maintenance practices?

Table 24.2. Basis for assessment of gains and losses due to maintenance from a sustainability perspective: some examples

Key issues for assessment of <i>financial impact</i> of maintenance
Operating expenses
Tied-up capital for tools and other resources
Actual production quality/volume <i>vs</i> set targets
Actual conversion costs <i>vs</i> target conversion costs
Product quality performance
Insurances, compensations, and penalties
Key issues for assessment of <i>social impact</i> of maintenance
Plant safety integrity level and safety performance
Physical working environment
Occupational health and hygiene
Hazardous exposure
Key issues for assessment of <i>environmental impact</i> of maintenance
Toxic emissions (CO ₂ , NOx, <i>etc.</i>)
Waste production
Energy consumption
Volume of rejections, scraps and re-works of product
Level of usage of chemicals and other environmentally challenging mediums for maintenance work

If a given industrial asset has to be managed ensuring a specific technical condition, then the maintenance process needs to contribute in all the three stages (see Figure 24.8).

Obviously, the design basis has a major impact on an industrial facility that needs to be operated in an acceptable technical condition. For many years researchers have argued that maintenance should be incorporated even at the conceptual engineering stage of an industrial complex to make them maintenance friendly during production/manufacturing/process operations. The integration of RAMS (reliability, availability, maintainability, supportability) issues has gained much attention in this regard lately. One of the advantages of early involvement of maintenance experts, in fact, is the ability to identify and define feasible technical solutions not only to accommodate ‘*design for maintenance*’ aspects but also to open-up and facilitate the adoption of promising technical solutions to ‘*design-out maintenance*’. Very often, maintenance is seen to be taking a relatively passive role during plant design processes, particularly due to financial constraints. During the operational phase, on the other hand, maintenance plays relatively a more active role as inspection and maintenance programs begin to get operationalized. However, it is public knowledge that asset maintenance process is encountered with unprecedented challenges when a plant/facility inherits technical problems by default due to design faults and failures. Some of the problems can be designed out

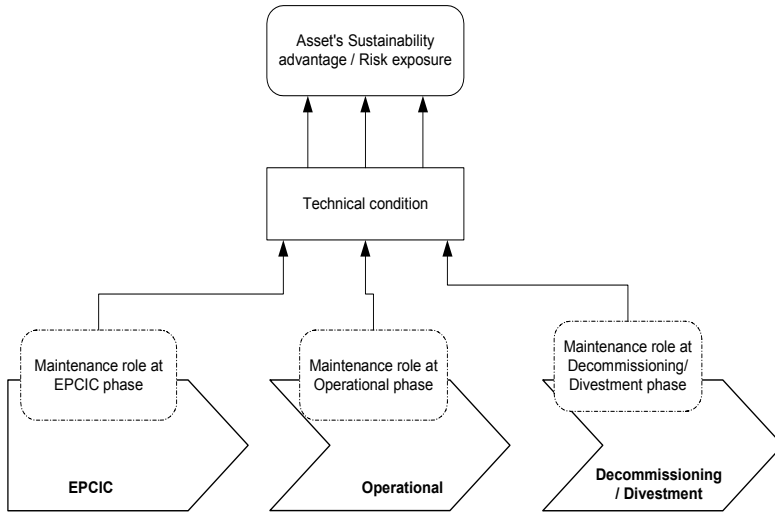


Figure 24.8. Illustration of maintenance *custodianship* in an asset's sustainability performance from a holistic life-cycle perspective

during the operational phase through appropriate modification efforts, while in most cases such modifications are known not to be commercially feasible due to large costs involved in making major technical changes to rectify problem areas. The decommissioning (or divestment) phase in most cases raises the need for a minimum maintenance policy. Either the statutory and regulatory requirement with respect to exposed risk of an end-of-life facility or the commercial interests of the owners to reuse the systems/equipment, or in fact the both, often set the basis for maintenance activities at the disposal/divestment phase.

The interesting question is what issues are important to consider in order to meet the necessary technical conditions that influence the sustainability performance of a given industrial facility. Table 24.3, for instance, illustrates different roles of maintenance during the three major stages of an asset's life.

In addition to the technical condition, another defining aspect of influence is the maintenance work quality. Certainly, the issue of maintenance quality has two-fold effects. First, it can have some direct impact on the availability and reliability performance (*i.e.*, technical condition). Second, it may also have some direct impact on some consequences independent of technical condition such as safety performance, amount of waste produced or environmental damage, excessive energy or resource consumption, *etc.* Work quality related issues, in this particular context, can be taken into account for example in terms of:

- Occupational negligence (for instance of safety precautions);
- Behavior and attitudes during work execution process;
- Voluntary procedural deviations; and
- Work priority manipulations/negligence, *etc.*

Table 24.3. Involvement of maintenance during the three major stages of an industrial facility to operate in an acceptable technical condition decisive for sustainability performance

EPCIC phase
Maintenance scenarios to manage future threats and opportunities
Defining maintenance related design basis to set acceptable standards for functional integrity
Identify and define feasible maintenance work philosophies and programs
Technical quality compliance strategy for third-party systems and equipment suppliers
Execution of risk and vulnerability analysis (including reliability, hazard and operability, maintainability and supportability, <i>etc.</i>)
Goal setting and responsibility charting
Document compliance and control process
Competence mapping and development procedures
Development of work processes and B2B organizational solutions
Damage proof storage and logistic solutions
<i>etc.</i>
Operational phase
Technical condition optimization with respect to plant performance targets
Continuous revision and update of maintenance philosophies and programs
Continuous update and effective management of technical documentation
Continuous integrity analysis and review of life-cycle costs
Analysis of performance trends and historical losses to map operational risk exposure
Competence revisions and management
Audits and verifications of inspection, testing, and maintenance activities
Continuous criticality analysis and work priority setting
<i>etc.</i>
Decommissioning/Divestment phase
Integrity and remaining useful life analysis
Condition assessment and re-usability analysis in the new operating set-up
Assessment of risk exposure
Removal and re-installation planning
<i>etc.</i>

These elaborations in principal are the foundational issues for expression of maintenance *custodianship* with respect to an asset’s sustainability performance. It mainly emphasizes that the emerging sustainability practice provides a stronger basis to challenge the traditional pure financial criteria for maintenance impact assessment. In the plant impact assessment process, through maintenance related *gains* and *loss* analysis, a broader set of aspects come into play with respect to an asset’s financial, social, and environmental performance criteria. In fact, an excellent contribution by maintenance towards a healthy and a commercially successful asset, in this regard, relies on both technical condition level (including technical safety integrity) of systems/equipment and the maintenance work quality. This is illustrated in Figure 24.9.

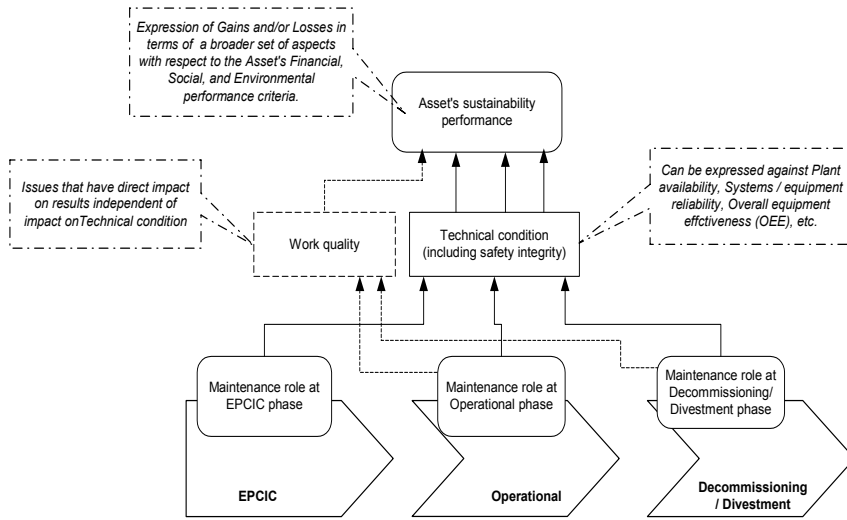


Figure 24.9. Foundational issues for expression of maintenance *custodianship* with respect to an asset's sustainability performance

Interestingly, the issues discussed in this section, and the generic frameworks illustrated, can be used in the development of a maintenance impact management process. Such impact management plays a vital role in so far as it can help support in identifying strengths and weakness in the maintenance process in corporate sustainability performance compliance efforts. Depending on the criticality of identified weaknesses, feasible solutions can be drawn to overcome those in such a way as to avoid negative effects either at work quality or technical condition levels. A preliminary structure of impact management process is discussed in the next section.

24.6 Generic Maintenance Impact Management Process

The engineering environment comprises different types or classes of industrial assets. Two principle classes include:

- Production/manufacturing / process assets; and
- Infrastructure assets.

The production/manufacturing/process assets are used in producing a product either discretely or continuously. Infrastructure assets on the other hand are represented by, for instance, electricity supply grids, gas distribution networks, logistics networks, *etc.* While each type of asset has a specific role in ensuring sustainability performance, what aspects are important and what needs to be prioritized can vary substantially. For example, there is a clear difference between a chemical processing plant and a hydroelectricity distribution grid, and thus between the underlying maintenance philosophies, policies, and practices that

would best work with each type of asset. It implies that the maintenance impact management process, for sustainability performance, can be very specific to a specific class of assets.

Furthermore, even in the same industry, not all the asset owners or operators have/aspire to similar sustainability related specifications to achieve commercial excellence. A business in this regard can be affected by number of factors such as:

- Share in the prospective markets;
- Globalization and business development plans;
- Competitive position within the representative industrial sector; and
- Growth strategies and preferred emerging markets, *etc.*

Moreover, asset owners or operators often have a number of assets in their portfolio that can be geographically distributed. This implies that the conditions under which a given asset has to be managed and the factors that influence the management of those assets can vary significantly. This happens due to unavoidable effects of:

- Design features (levels of complexity, technology in use, *etc.*);
- Age of the facility and the characteristics of the asset support system;
- Statutory and regulatory requirements;
- Technical infrastructures critical to daily operations; and
- Socio-economic and socio-technical conditions, *etc.*

Such business needs, expectations, and/or requirements, coupled with the operational conditions that an asset has to undergo, have a significant impact on the definition and development of suitable sustainability management frameworks. Very often, such frameworks developed by businesses for their executive leadership remain very generic and abstract. It is defined in a way that it applies to all sections of the business and the entire portfolio of assets, regardless of specific characteristics. At the individual asset level, the asset manager and/or operational manager needs to take responsibility for appropriately transforming business issues into guiding managerial frameworks for daily management of assets. This introduces a demand basis for distinctive asset processes such as maintenance, production, quality control, *etc.*

In fact, maintenance impact management is a process quite specific to a given asset under particular business requirements and operational conditions. Figure 24.10 outlines a generic set-up for the maintenance impact management process that can be customized depending on the asset class in question.

The impact management process can be influenced, as shown in Figure 24.10, by various internal performance constraints, for instance:

- Ageing workforce and inadequate competence availability;
- Budgetary limitations and inventory policies;
- Obsolescence of spares and other material resources; and
- Logistics challenges, *etc.*

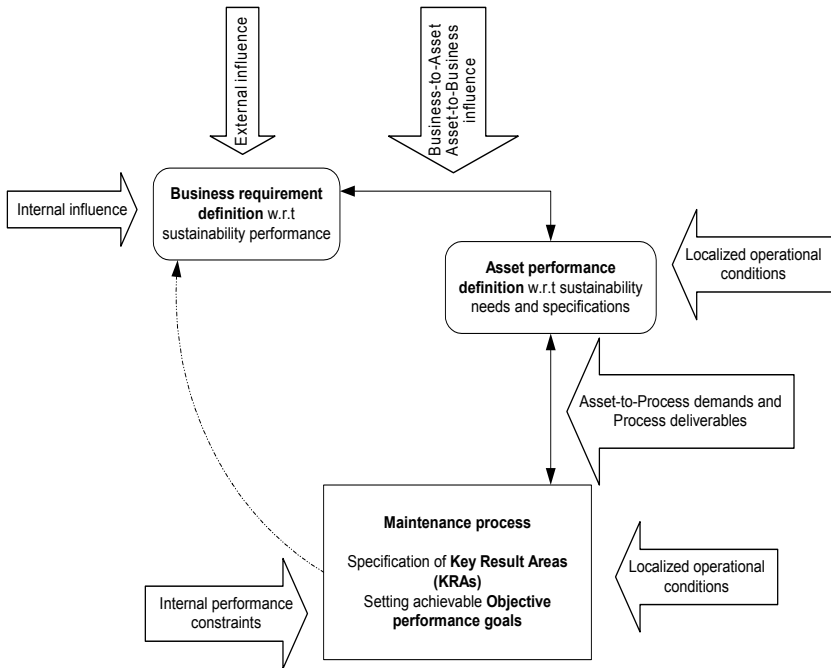


Figure 24.10. Generic structure for maintenance impact management process

Under such constraints and operational conditions, a key challenge for maintenance managers is to identify and define clearly *Key Result Areas* (so-called *KRAs*). This is perhaps the most critical task of the impact management process, since they set the point of departure for defining *Critical Success Factors (CSFs)* and *Key Performance Indicators (KPIs)* that help support achieving *objective performance goals* related to each KRA. If the set of KRAs is ill-defined, then the potential to resort to ill-defined or poor CSFs is relatively large resulting in insignificant impact or even failing to meet performance expectations.

This implies that KRAs play a pivotal role for maintenance (or for any asset process for that matter) in the impact management process. Some exemplary KRAs are given in Table 24.2. The specification and analysis window for maintenance impact management process is illustrated in Figure 24.11.

Some specific examples related to *specific conditions* and *influence factors* (focusing on commercial life-cycle) were highlighted in the previous section.

While CSFs and KPIs have major roles in the maintenance impact management process, two other integral components are also important within the same context, *i.e.*:

- Quality assurance; and
- Best practice implementation.

Quality assurance processes aim at ensuring better control and coordination of CSFs, and the compliance to operational standards. Best practice implementation

efforts, on the other hand, need to aim at the use of most effective techniques and methods to meet objective performance goals.

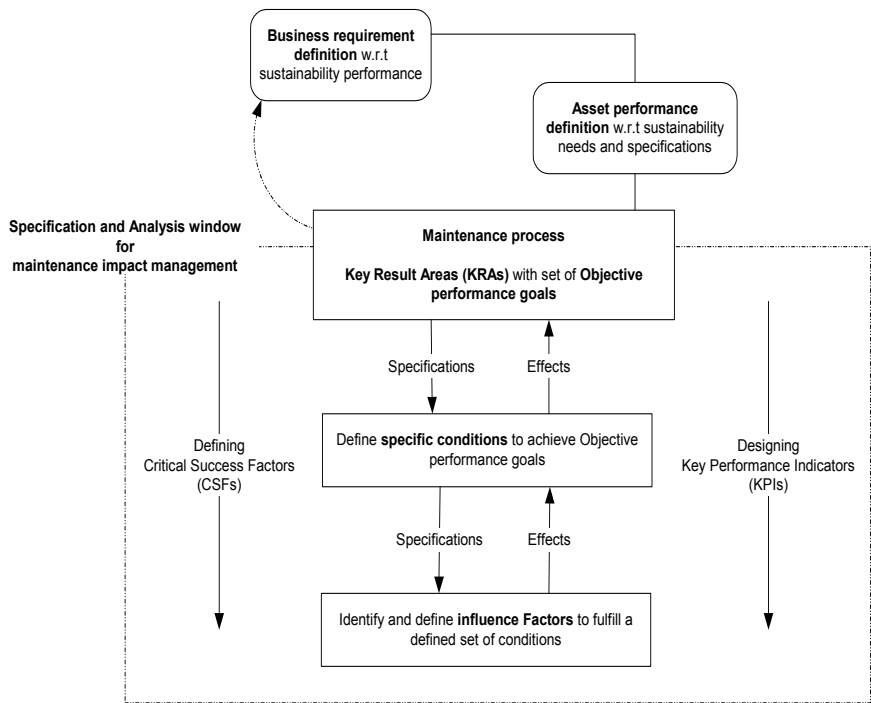


Figure 24.11. Specification and analysis window for maintenance impact management process

24.7 Adapting an Effective Asset Maintenance Practice for Sustainability

The changes in the criteria used to evaluate risk-based decisions are apparent, moving from a pure economic perspective historically to the consideration of social and environmental factors (Wolff *et al.* 2000). This entails that asset maintenance practices, too, must be comprehensive to encompass financial performance assessment as well as environmental protection and societal impact (Liyanage, 2003; Liyanage and Kumar, 2003).

Adopting a sustainability-oriented holistic view to asset maintenance management is pivotal to successful business operations. The application of novel and innovative approaches can open up significant opportunities for businesses, particularly in capital-intensive sectors such as oil and gas exploration. This also has great potential to improve core business performance in environments with increased sustainability risk due to tighter regulatory control and increased social awareness demanding ethical and responsible corporate practices.

Quality of asset maintenance process plays a vital role in sustainability compliance efforts. Quality assurance to enhance the technical condition and performance of industrial assets inherently requires focus on the asset's entire life cycle, from the EPCIC phase through operational and decommissioning/divestment phases, to ensure superior performance through an effective maintenance process. Given the increased competitiveness in markets, companies have to utilize maintenance strategies that would minimize the risk and uncertainty in business operations. Thus, mitigating the sustainability risk through asset maintenance, which no longer is a mere support function, postulates the integration of the sustainability compliance measures across all three phases of the assets' life to enhance asset productivity, reliability, and safety, and thereby increase asset value in the core business. In the context of the maintenance process, quality assurance signifies better control and coordination of the CSF or performance drivers. The CSF that entail superior performance with respect to the three pillars of sustainability needs to be defined first followed by identifying KPI to enable measurement and evaluation of these CSF to ensure the desired performance criteria have been achieved.

Industrial assets designed with emphasis on abating maintenance requirements, and prolonging asset usage during subsequent life cycle phases, are favored by asset users as well as manufacturers. Many approaches have traditionally been limited, largely, to life cycle cost analysis to minimize the overall costs of maintenance (Dwight, 1995; Wolff *et al.* 2000) with minimal emphasis on other broader issues. Most of the asset maintenance needs are determined during the EPCIC phase, particularly during design (Blanchard and Fabrycky, 1998). Ensuring a sustainability-focus, and quality assurance, in asset maintenance thus necessitates an enhanced emphasis on environmental as well as social implications of maintenance during the operational phase while the assets are being designed. On the other hand, the operational phase being the most economically crucial in the productive life of an asset, some efforts have been devoted to enhance asset maintenance quality assurance during this stage as well (see for example Sherwin, 2000; Schneider *et al.* 2006; Pramod *et al.* 2006 for reviews).

Sustainability-orientation in maintenance focus during the decommissioning and divestment phase necessitates the consideration of the 6R approach of *reduce, reuse, recycle, recover, redesign, and remanufacture* (Jawahir and Dillon, 2007). Extending the use of assets through multiple life cycles aimed at 'near perpetual' material usage (Jaafar *et al.* 2007) enables prolonged use of scarce resources, to reduce adverse environmental impacts and increase productivity. Conventionally, the emphasis has been on decommissioning assets to minimize interference with production and processing operations, mitigating only financial implications; this needs to be extended to consider the potential of recovering and reusing the assets, or parts therein, through redesign and remanufacturing as is already being done by many companies, particularly in the industrial machinery sector (Ferrer and Whybark, 2001). Further, the isolated consideration of the decommissioning/divestment phase makes achieving quality in maintenance during this stage more difficult. Rather, assets must be designed and commissioned, during the EPCIC phase, to facilitate application of 6R concept during the last phase of its life to achieve overall sustainability.

Best practices for asset maintenance can be derived by benchmarking the broad range of tools, techniques, methods and strategies applied by successful companies in achieving the standards of excellence. With the emergent sustainability focus, best practices to create value in asset maintenance (*i.e.*, the value-based concept), too, are continuously evolving. Technology has been a major change agent in practically all industrial sectors, affecting asset design, functioning as well as maintenance operations (Tsang *et al.* 1999). Environmental regulations are also continuously changing with governments introducing new legislations to mitigate impacts due to adverse corporate activities (Holliday *et al.*, 2002). Further, societal expectations have been evolving, with increased pressure on organizations to adopt better health and safety standards and provide better quality of work life (WBCSD, 2006). All these enumerate that best practices in asset maintenance to mitigate sustainability risk need to evolve continuously, throughout all the life cycle phases of the assets. This means that companies need to engage in a continuous improvement process, following the classical plan, do, check, act (PDCA) philosophy advocated by Deming (1982), to benchmark internally between different business processes as well as externally between various organizations, and even industries, to achieve best practice in terms of economical, environmental and social sustainability in asset maintenance across all life cycle phases, as indicated in Figure 24.12.

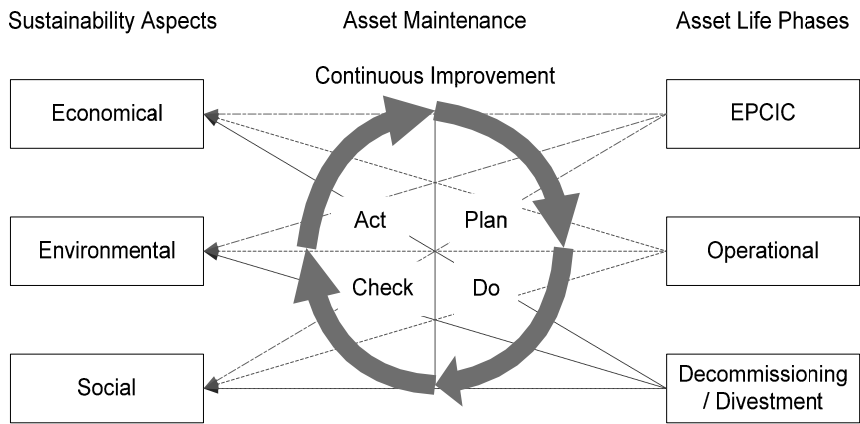


Figure 24.12. Continuous improvement in asset maintenance for sustainability

Emerging concepts in sustainable innovation, to increase value creation, show much potential in quality assurance and promoting better practices for sustainable maintenance operations in the future. One such concept is product-service systems (PSS) (Mont and Pelpys, 2003), which shift the focus from “*designing and selling physical products only, to selling a system of products and services...*” (Manzini and Vezzoli, 2001). PSS sell a utility to consumers, leaving product ownership, maintenance, parts recycling and eventual product decommissioning and replacement in the hands of manufacturers (Manzini and Vezzoli, 2001; Piller, 2003), significantly increasing opportunities to incorporate the 6R concept for

sustainable product design and manufacturing to improve maintenance performance. PSS is an emerging concept and research on promoting sustainable maintenance operations through PSS is still in its infancy or literally non-existent and need to be fully studied.

Also recently the potential of applying the Balanced Scorecard (BSC) model as a basis for continuous improvement in maintenance management has been discussed (Tsang *et al.* 1999; Liyanage, 2003). According to Liyanage (2003), though the conventional BSC model has a more economic emphasis (as also discussed by Neely and Adams, 2001), the perspectives considered in the original model can be mapped to assess the impact in terms of economical, environmental and societal concerns. Such efforts can help guide any efforts deployed towards quality assurance, best practice implementation, continuous improvement, and performance standardization. This encompasses the idea, as presented by Liyanage (2003), that attaining the sustainability goals requires using strategic resources and core competencies to determine the capabilities needed to maintain asset condition and achieve desired results to meet the triple bottom-line of sustainability.

24.8 Conclusion

Industrial maintenance has long been viewed as a cost center within plant/facility management processes. As various interest groups gradually began to realize and echo the societal and environmental implications of commercial activities, the concept of sustainable business practice gathered momentum. Today, discussions and debates continue in many corners of the socio-political and socio-economical environment about the significance of sustainability thinking and practices for the betterment of the common world. This chapter presented a discussion about the asset maintenance process in light of the emerging sustainability concerns. The purpose is to portray the value-added nature of asset maintenance as it has emerged through sustainability thinking and to challenge the conventional financially-oriented view of maintenance. It implies that maintenance processes in fact have significant contributions to businesses in terms of economical, environmental, and societal implications of the commercial activities.

References

- Banks RI and Wheelwright SC (1979) Operations versus strategy: trading tomorrow for today. *Harvard Business Review*, May-June.
- Blanchard BS (2001) Maintenance and Support: A Critical Element in the System Life Cycle, Proceedings of the International Conference on Maintenance Societies, Melbourne, Australia Paper 003.
- Blanchard BS and Fabrycky WJ (1998) *Systems Engineering and Analysis*, 3rd Edition, Prentice-Hall, NJ.
- Briassoulis H (2001) Sustainable Development and its indicators: Through a (planner's) glass darkly. *J of Environ Plan and Manag* 44(3): 409–427.
- Caldwell LK (1989) A Constitutional Law for the Environment: 20 years with NEPA indicates the need. *Environ* 31(10): 6–11, 25–28.

- CorpWatch (2008) <http://www.corpwatch.org/article.php?id=1348#definitions>. Accessed April 4, 2008.
- Committee on Industrial Environmental Performance Metrics (Author), National Academy of Engineering and National Research Council (Author), 1999. Industrial environmental performance metrics: challenges and opportunities. Washington, DC, National Academy Press, August 24th pp. 252.
- Cross KF and Lynch RL 1988–1989 The SMART way to define and sustain success. *Natl Prod Rev* 9(1): 23–33.
- Daly HE (1990) Toward some operational principles of sustainable development. *Ecol Econ* 2: 1–6.
- Daniel JT (2005) Corporate Social Responsibility: A Comparison Between the United States and the European Union. *Corpo Soc Responsib and Environ Manag* 12(1): 55–59.
- Deloitte andTouche, ISSD (1992), Business strategy for sustainable development: leadership and accountability for the 90s. <http://www.iisd.org/publications/pub.aspx?id=242>
- Deming WE (1982) Quality, Productivity and Competitive Position, MIT Press, Cambridge, MA.
- Dixon F (1999) Environmental Leaders Achieve Superior Stock Market Performance in the ElectricUtility Sector. The Annual Public Utility Reporters Environmental Conference, New Orleans, LA. May 25.
- Doering DS, Cassara A, Layke C, Ranganathan J, Revenga C, Junstall D and Vanasselt W (2002) Tomorrow's markets: Global trends and their implications for business, collaboration of World Resources Institute, United Nations Environment Programme, and World Business Council for Sustainable Development.
- Dormann J and Holliday C (2002) Innovation, Technology, Sustainability and Society, World Business Center for Sustainable Development.
- Dwight RA (1995) Concepts for Measuring Maitnenace Performance, New Developments in Maintenance. Moret Ernst & Young Management Consultants, Netherlands, pp. 109–125.
- Elkington J (1998) Cannibals with Forks: The Triple Bottom Line of the 21st Century Business, New Society Publishers, USA.
- Elliott SR(2005), Sustainability: An economic perspective. *Resour Conserv and Recycl* 44(3): 263–277.
- Epstein MJ (1996) Measuring Corporate Environmental Performance: Best Practices for Costing and Managing an Effective Environmental Strategy, Institute of Management Accountants, Irwin, Chicago, IL.
- Ferrer G and Whybark DC (2001), Material Planning for a Remanufacturing Facility. *Prod and Oper Manag* 10(2): 112–124.
- Ghalayini AM and Noble JS (1996) The changing basis of performance measurement. *Int J of Oper & Prod Manag* 16(8): 63–80.
- Hart SJ (1997) Beyond Greening: Strategies for a Sustainable World. *Harv Bus Rev*. January/February.
- Hayes RH and Garvin DA (1982) Managing as if tomorrow mattered. *Harv Bus Rev*. 60(3): 70–79.
- Holliday C and Peppers J (2001) Sustainability through the markets – Seven keys to success, World Business Council on Sustainable Development, Geneva.
- Holmberg J and Sandbrook R (1992) Sustainable Development: What to be done? In: Holmberg, J. (ed), *Politics for a Small Planet*, Earthscan, London.
- Holliday CO, Schmidheiny S, and Watts P (2002) Walking the talk: The business case for sustainable development, Greenleaf Publishing, Sheffield, U.K.
- Jaafar IH, Venkatachalam A, Joshi K, Ungureanu AC, De Silva N, Dillon OW Jr, Rouch KE, and Jawahir IS (2007) Product Design for Sustainability: A New Assessment

- Methodology and Case Studies. Handbook of Environmentally Conscious Mechanical Design, (Editor: M. Kutz), John Wiley & Sons, Chapter 2, pp. 25–65.
- Jawahir IS and Dillon OW (2007) Sustainable Manufacturing Processes: New Challenges for Developing Predictive Models and Optimization Techniques, (Keynote Paper), Proc. 1st International Conference on Sustainable Manufacturing (SM1), Montreal, Canada, October 18–19, pp. 1–19.
- Jawahir IS and Wanigaratne PC (2004) New challenges in developing science-based sustainability principles for next generation product design and manufacture. Proceedings of the 8th International Research/Expert Conference on Trends in the development of machinery and associated technology, pp 1–10.
- Johnson HT and Kalpan RS (1987) *Relevance Lost: The Rise and Fall of Management Accounting*. Cambridge, MA: Harvard Business School Press.
- Kaplan RS (1983) Measuring manufacturing performance: a new challenge for managerial accounting research. *Account Rev* 58(4): 686–703.
- Kaplan RS (1984) Yesterday's accounting undermines production. *Harv Bus Rev* 62: 95–101.
- Kaplan RS and Norton DP (1996) *The Balance Scorecard*, Harvard Business School Press, Boston, MA.
- Keeble JJ, Topiol S, and Berkeley S (2003) Using indicators to measure sustainability performance at a corporate and project level. *J Bus Ethics*, 44: 149–158.
- Keijzers G (2002) The Transition to the Sustainable Enterprise. *J Clean Prod* 10: 349–359.
- Khadem R (1988) One page management: a unified approach to productivity. *Natl Prod Rev* 8(1): 45–57.
- Kiernan MJ and Martin JS (1998) Wake-Up Call for Fiduciaries: Eco-Efficiency Drives Shareholder Value, *Today's Corporate Investor*, December, pp. 17–18.
- Liyanage JP (2003) Operations and maintenance performance in oil and gas production assets: Theoretical architecture and Capital value theory in perspective, PhD thesis, Norwegian University of Science and Technology (NTNU), Norway.
- Liyanage JP (2007) Operations and maintenance performance in production and manufacturing assets: The sustainability perspective. *J of Manuf Technol Manag* 18(3–4): 304–314.
- Liyanage JP and Kumar U (2003) Towards a value based view on operations and maintenance performance management. *J of Qual in Maint Eng* 9(4): 333–350.
- Manzini E and Vezzoli C (2001) Product Service Systems and Sustainability: Opportunities for Sustainable Solutions, United Nations Environment Program.
<http://www.uneptie.org/pc/sustain/reports/pss/pss-imp-7.pdf>
- Markeset T (2003) Dimensioning of product support: Issues, challenges, and opportunities, PhD thesis, Stavanger University College, Norway.
- McNair CJ and Mosconi W (1987) Measuring performance in an advanced manufacturing environment. *Manag Account* 69(1): 28–31.
- Mont O and Pelpys A (2003) Customer Satisfaction: Review of Literature and Application to the Product-Service Systems. International Institute for Industrial Environmental Economics, Lund University, Sweden.
- Muthu S, Devadason SR, Ahmed S, Suresh P and Baladhanadayutham, (2000), Benchmarking for Strategic Maintenance Quality Improvement. *Benchmarking. An Int J* 7(4): 292–303.
- Neely A and Adams C (2001) Perspectives on Performance; The Performance Prism, Focus Magazine for the Performance Management Professional, Issue No. 4, E-magazine available at http://www.focusmag.com/back_issues/issue_04/index.htm, accessed on July 5, 2007.
- Oldenburg KU and Geiser K (1997) Pollution prevention and...or industrial ecology. *J of Clean Prod* 5(1-2): 103–108.

- Piller, F.T. (2003), Turning Products into Sustainable Service Businesses by Mass Customization, Presentation at the Sustainable Product-Service-Systems: State of the art, Launch Conference of the EU-funded Thematic Network on Sustainable Product Service Systems (SusProNet), June, 2006
- Pramod VR, Devadason SR, Muthu S, Jagathiraj VP and Dhakshina Moorthy G, (2006) Methodology and Theory: Integrating TPM and QFD for Improving Quality in Maintenance Engineering. *J of Qual in Maint Eng* 12(2): 150–171.
- Ranganathan J (1998) Sustainability Rulers: Measuring Corporate Environmental & Social Performance. *Sustainable Enterprises Perspectives*. Washington DC, World Resource Institute.
- Ratnayake RMC and Liyanage JP (2007) Corporate dynamics vs. Industrial asset performance: The sustainability challenge, The 2nd World Congress on Engineering Asset Management (WCEAM 2007), Harrogate, UK., pp 1645–1656.
- Santori PR and Anderson AD (1987) Manufacturing performance in the 1990s: measuring for excellence. *J of Account* 164(5): 141–147.
- Schmidheiny S (1992) Changing Course: A Global Business Perspective on Development and the Environment, the MIT Press.
- Schneider J, Gaul AJ, Neumann C, Hografer J, WellBow W, Schwan N, and Schnettler A (2006) Asset Management Techniques. *Electr Power and Energy Syst* 28: 643–654.
- Sherwin D (2000) A Review of Overall Models for Maintenance Management. *J of Qual in Maint Eng* 6(3): 138–164.
- Tsang AHC (2002) Strategic Dimensions of Maintenance Management. *J of Qual in Maint Eng* 8(1): 7–39.
- Tsang AHC and Brown WL (1998) Managing the Maintenance Performance of an Electric Utility through the use of Balance Scorecards, Proceedings of the 3rd International Conference of Maintenance Societies, Adelaide, Australia, pp. 1–10.
- Tsang AHC Jardine AKS, and Kolodny H (1999) Measuring Maintenance Performance: A Holistic Approach. *Int J of Oper & Prod Manag* 19(7): 691–715.
- UN World Commission on Environment and Development (1987) Our Common Future.
- Van Dieren, W., (1995), Taking Nature into Account: Toward a Sustainable National Income. A Report to the Club of Rome. Copernicus, Springer-Verlag. New York.
- Warhurst A (2002) Sustainability Indicators and Sustainability Performance. Management. Mining and Energy Research Network, Report No 43, Warwick Business School, with IIED and WBCSD. University of Warwick.
- WBCSD (2004) Running the Risk – Risk and Sustainable Development: a Business Perspective, World Business Council for Sustainable Development
- WBCSD (2006) From Challenge to Opportunity: The Role of Business in Tomorrow's Society, World Business Council for Sustainable Development.
- Wilson M (2003) Corporate sustainability: What is it and where does it come from? *Ivey Business Journal*, (March/April), pp. 1–5.
- Wireman T (2004) Benchmarking Best Practices in Maintenance Management, Industrial Press Inc., New York, NY.
- Wolff R, Zaring O, Furuhoft E, O'Leary H, Gaasbeek AP and Thoem T (2000) Indicators for Sustainable Development, Paper presented at the The SPE International Conference on Health, Safety, and the Environment in Oil and Gas Exploration and Production, Stavanger, Norway, June 26–28.
- World Business Council on Sustainable Development (WBCSD) (2002) The Business Case for Sustainable Development. Making a difference toward the Johannesburg Summit 2002 and beyond. Geneva.
- Young R (2006) Sustainability: From rhetoric to reality through markets. *J of Clean Prod* 14: 1443–1447.

Zadek S, Pruzan P and Evans R (1997) Building Corporate Accountability. Emerging practices in social and Ethical Accounting, Auditing and Reporting. Earthscan. London, UK.

Human Reliability and Error in Maintenance

B.S. Dhillon

25.1 Introduction

Although humans have felt the need for maintenance of their equipment since the beginning of time, the beginning of the modern engineering maintenance may be regarded as the development of steam engine by James Watt [1736 – 1819] in 1769 in Great Britain (The Volume Library, 1993). Today, billions of dollars are being spent each year on equipment maintenance around the world. For example, each year United States industry alone spends over \$300 billion on plant maintenance and operations and for the fiscal year 1997, the operation and maintenance budget request of the United States Department of Defense was \$79 billion (Latino, 1999; 1977 DoD Budget, 1996).

Humans play an important role during equipment life cycle: design, production, and operation and maintenance phases. Even though, the degree of their role may vary from one equipment to another and from one equipment phase to another, it is subject to deterioration because of the occurrence of human error. A human error may be classified under six distinct categories: design, assembly, inspection, installation, operating, and maintenance (Meister, 1962, 1976). In particular, the maintenance error or poor human reliability occurs basically because of wrong repair or preventive measures and their two examples are incorrect calibration of equipment and application of the wrong grease at appropriate points of the equipment. A comprehensive list of publications on human reliability and error in engineering maintenance is available in Dhillon and Liu (2006). This chapter presents various important aspects of human reliability and error in maintenance.

25.2 Terms and Definitions

This section presents terms and definitions considered, directly or indirectly, useful for studying human reliability and error in maintenance (MIL-STD-721B, 1966; Dhillon, 1986, 2002; Hagen, 1976; AMCP, 1975; McKenna and Oliverson, 1997; Omdahl, 1988; Naresky, 1970):

- *Human error*: the failure to perform a specified task (or the performance of a forbidden action) that could result in disruption of scheduled operations or result in damage to equipment and property;

- *Human reliability*: the probability of accomplishing a task successfully by humans at any required stage in system operation within a stated minimum time limit (if the time requirement is specified);
- *Maintenance*: all actions necessary for retaining an item or equipment in, or restoring it to, a specified condition;
- *Inspection*: this is the qualitative observation of condition or performance of an item;
- *Human performance*: a measure of man-functions and actions under specified conditions;
- *Man-function*: that function which is allocated to the system's human element;
- *Corrective maintenance*: the unscheduled maintenance or repair actions to put back items/equipment to a specified state and performed because maintenance personnel or users perceived failures or deficiencies;
- *Predictive maintenance*: the use of modern measurement and signal-processing approaches to diagnose equipment condition during operation;
- *Human performance reliability*: the probability that a human will satisfy all stated human functions subject to specified conditions;
- *Continuous task*: a task that involves some kind of tracking activity; and
- *Preventive maintenance*: all actions performed on a planned, periodic and specific schedule for keeping an item or equipment in specified working condition through the process of reconditioning and checking.

25.3 Human Reliability and Error in Maintenance-Related Facts, Figures, and Examples

Some of the important facts, figures, and examples directly or indirectly associated with engineering maintenance are as follows:

- Each year, the United States industry spends over \$300 billion on plant maintenance and operations (Latino, 1999).
- A study of 213 maintenance events reported that 25.8% of the failures were partially or wholly due to human error (Robinson *et al.* 1970).
- A study of safety issues vs on board fatality of worldwide fleet of jets of the period 1982 to 1991, revealed that inspection and maintenance was the second most pressing safety issue in regard to onboard fatalities of 1481 (Russell, 1994; BASI 1997).
- In 1993, a study of 122 maintenance-related occurrences involving human factors revealed that the classifications of maintenance error breakdowns were omissions (56%), wrong installations (30%), wrong parts (8%), and other (6%) (BASI, 1997; Circular 243, 1995).
- A study of electronic equipment concluded that approximately 30% of all malfunctions were the result of operation and maintenance errors (AMCP, 1972).

- A study of tasks such as align, adjust, and remove concluded a human reliability mean of 0.9871 (Sauer *et al.* 1976).
- A study of maintenance operations among commercial airlines concluded that 40–50% of the time non-defective parts were removed for repair (Christensen and Howard, 1981).
- A study of maintenance errors in missile operations revealed many different causes: dials and controls (miss-read, miss-set) (38%), wrong installation (28%), loose nuts/fittings (14%), inaccessibility (3%), and miscellaneous (17%) (Dhillon, 1986; Christensen and Howard, 1981).
- In 1979, at O'Hare airport in Chicago, in a DC-10 aircraft accident, a total of 272 people died because of incorrect procedures followed by the maintenance personnel (Tripp, 1999).
- In 1983, an L-1011 aircraft departing Miami, Florida lost oil pressure in all its three engines because of missing chip detector O-rings. A subsequent investigation traced the problem to poor inspection and supply procedures (Dhillon, 1986; Tripp, 1999).
- An incident at Ekofisk oil field in the North Sea, involving the blow out-preventer (assembly of valves) was caused by inadvertent upside down installation of the device. The cost of the incident was estimated to be around \$50 million (Dhillon, 1986; Christensen and Howard, 1981).

25.4 Occupational Stressors, Human Performance Effectiveness, and Human Performance Reliability Function

There are many occupational stressors and they may be classified under four categories as shown in Figure 25.1 (Beech *et al.* 1982). The Category I stressors are concerned with problems pertaining to work load (*i.e.*, work overload or work under load). In the case of work overload the job requirements exceed the ability to satisfy them effectively. Similarly, in the case of work under load the work carried out by the individual does not provide meaningful stimulation. Some examples of work under load are repetitive performance, lack of opportunity to use one's acquired skills and expertise, and lack of any intellectual input.

The Category II stressors are concerned with problems pertaining to occupational frustration. More specifically, these problems lead to conditions where the job inhibits the meeting of set goals or objectives. Some of the factors that form elements of the occupational frustration are role ambiguity, lack of communication, bureaucracy difficulties, and poor career development guidance.

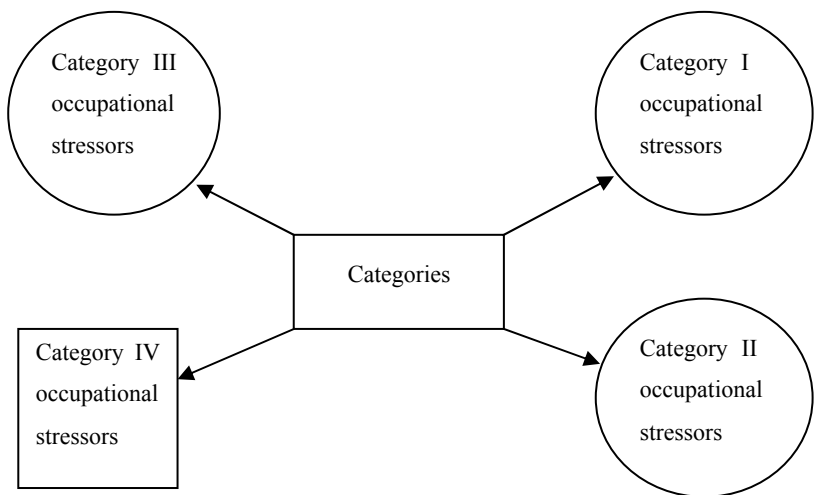


Figure 25.1. Four categories of occupational stressors

The Category III stressors are concerned with occupational change that disrupts one’s physiological, behavioural, and cognitive patterns of functioning. Some of the forms of occupational change are promotion, organizational restructuring, and relocation. All in all, these types of stressors are normally present in an organization concerned with productivity and growth. The Category IV stressors are all those stressors not included in Categories I, II, and III. Some examples of the possible sources of such stressors are too little or too much lighting, noise, and poor interpersonal relationships.

Over the years researchers have studied the relationships between stress and human performance effectiveness and have concluded the relationship between human performance effectiveness and stress as depicted by the curve shown in Figure 25.2 (Hagen, 1976; Beech *et al.* 1982). The curve shows that stress is not an entirely negative state. In fact, stress at a moderate level is necessary to increase human effectiveness to its optimum level. Otherwise, at very low stress the task will be dull and unchallenging; consequently, human performance will not be at its maximum.

In contrast, stress above a moderate level will cause human performance to decline due to factors such as fear, worry, and other kinds of psychological stress. All in all, the moderate stress may simply be defined as the level of stress sufficient to keep humans alert.

From time to time humans carry out various types of time-continuous tasks including aircraft manoeuvring, scope monitoring, and missile countdown. In performing tasks such as these, human performance reliability is an important parameter to consider. A general human performance reliability function or equation for time-continuous tasks can be developed in a similar way to the development of the general reliability function for hardware systems. Thus, we write the following equation for time dependent human error rate (Dhillon, 1986; Shooman, 1968):

$$\lambda_{he}(t) = -\frac{1}{HR(t)} \cdot \frac{dHR(t)}{dt} \quad (25.1)$$

where $\lambda_{he}(t)$ is the time t dependent human error rate. It is equivalent to time dependent failure rate of hardware systems. $HR(t)$ is the human performance reliability at time t .

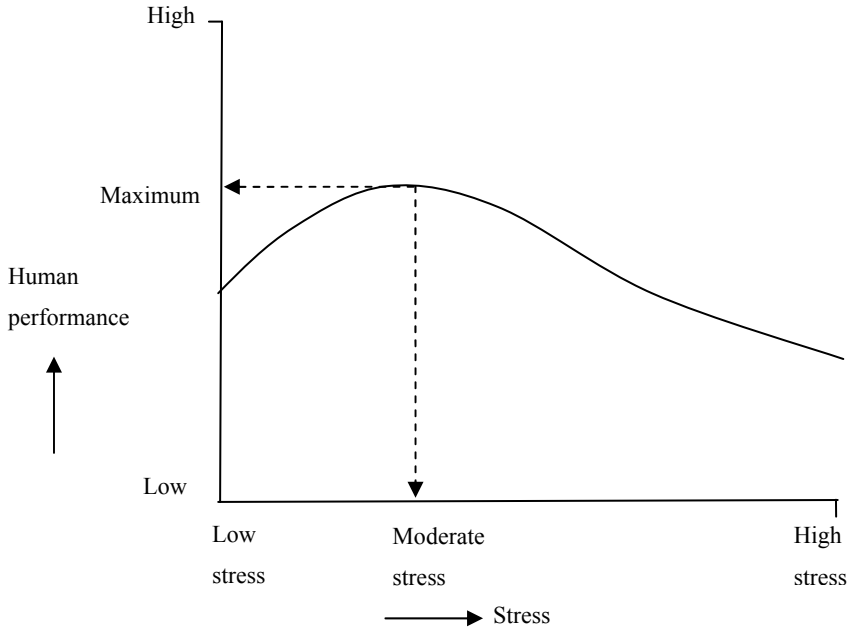


Figure 25.2. Human performance effectiveness vs stress relationship curve

By rearranging Equation 25.1 and then integrating both sides over the time interval $[0, t]$, we get

$$\int_0^t \frac{1}{HR(t)} \cdot dHR(t) = - \int_0^t \lambda_{he}(t) dt \quad (25.2)$$

Since at $t = 0$, $HR(t) = 1$, we rewrite Equation 25.2 as follows:

$$\int_1^{HR(t)} \frac{1}{HR(t)} \cdot dHR(t) = - \int_0^t \lambda_{he}(t) dt \quad (25.3)$$

After evaluating the left-hand side of Equation 25.3, we get

$$\ln HR(t) = - \int_0^t \lambda_{he}(t) dt \quad (25.4)$$

Thus, from Equation 25.4, we obtain:

$$HR(t) = e^{- \int_0^t \lambda_{he}(t) dt} \quad (25.5)$$

Equation 25.5 is the general human performance reliability function. It is applicable to both constant and non-constant human error rates. More specifically, Equation 25.5 is applicable when times to human error are described by probability distributions such as exponential, Weibull, gamma, and lognormal (Dhillon, 1986; Regulinski and Askren, 1969).

25.5 Human Error Occurrence Ways, Consequences, and Classifications, and Maintenance Error in System Life Cycle

There are many different ways in which human error can occur. The five most widely accepted ways are as follows (Hammer, 1980):

- *Way I:* performing a task that should not be performed;
- *Way II:* making an incorrect decision in response to a problem or difficulty;
- *Way III:* failure to recognize a hazardous situation;
- *Way IV:* failure to carry out a stated function; and
- *Way V:* poor timing and ineffective response to a contingency.

Useful guidelines to reduce the occurrence of human error in maintenance are presented in Section 25.8.

The consequence of human errors can vary quite significantly from one set of equipment to another or one task to another. In addition, a consequence may range from minor to severe. Nonetheless, the consequence of a human error in regard to a piece of equipment may be classified under three categories as shown in Figure 25.3 (Meister, 1962 and Dhillon, 1986).

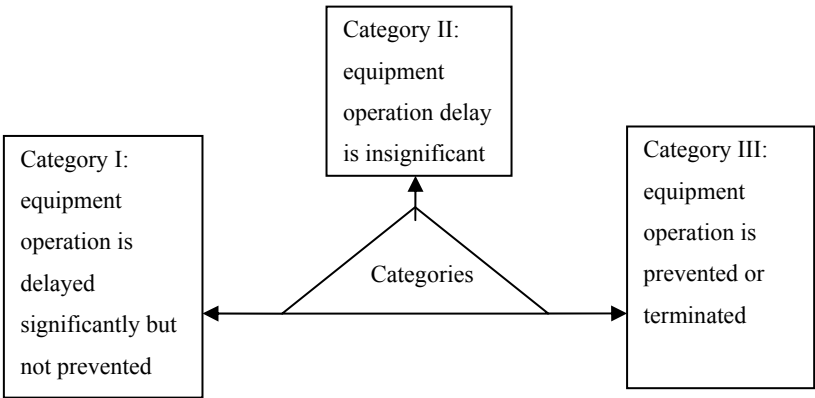


Figure 25.3. Categories of human error consequences in regard to equipment

Human errors may be broken down under various classifications. Six commonly used classifications in the industrial sector are as follows (Meister, 1962; Dhillon, 1986):Design errors;

- Inspection errors;
- Assembly errors;

- Installation errors;
- Operator errors; and
- Maintenance errors.

The occurrence of maintenance error in the system life cycle (*i.e.*, from the time of system acceptance to the beginning of its phase-out period) is an important factor. Approximate breakdowns of human errors (*i.e.*, assembly error, installation error, operator error, and maintenance error) that cause system failure in the system life cycle are shown in Figure 25.4 (Christensen and Howard, 1981; Hammer, 1980). The figure shows that the contribution of the maintenance error to the total human error is at least equal to that of operator error. Also, it is to be noted from the figure that as the system ages, the maintenance error increases quite dramatically.

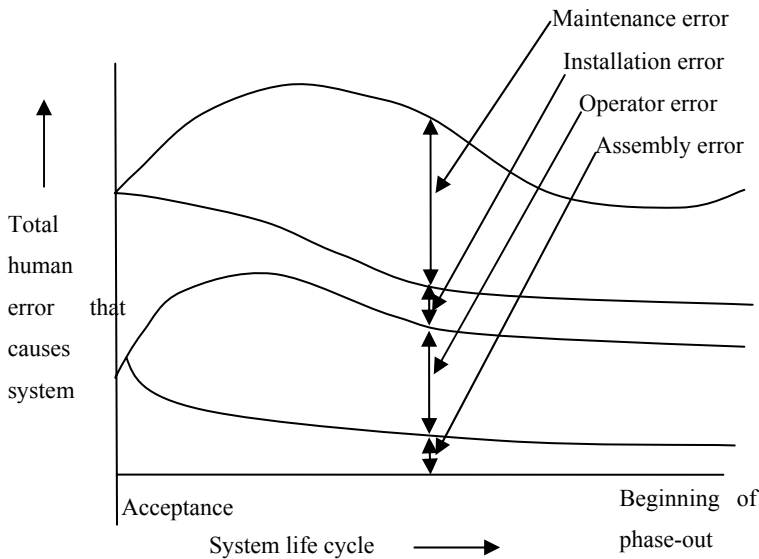


Figure 25.4. System life cycle vs four categories of human errors

25.6 Reasons for the Occurrence of Human Error in Maintenance and Top Human Problems in Maintenance

There are various reasons for the occurrence of human error in maintenance. Some of these are listed below (Dhillon, 1986; Christensen and Howard, 1981).

- Complex maintenance task;
- Inadequate or improper work tools;
- Poor equipment design;
- Poorly written maintenance procedures;

- Poor work layout;
- Outdated maintenance manuals;
- Fatigued maintenance personnel;
- Poor job environment (*e.g.*, lighting, humidity, and temperature); and
- Inadequate training and experience.

In particular, with regard to training and experience, a study of maintenance personnel concluded that those who ranked highest possessed characteristics as follows (Sauer *et al.* 1976; Christensen and Howard, 1981):

- More experience;
- Greater satisfaction with the work group;
- Greater emotional stability;
- Fewer reports of fatigue; and
- Higher aptitude and morale.

In addition, a study of correlation analysis revealed a significant degree of positive correlations between task performance and factors such as morale, years of experience, responsibility-handling ability, and amount of time in career field. The study also revealed a significant degree of negative correlations between task performance and anxiety level and fatigue symptoms.

Over the years, various studies have been performed to identify human factors-related problems in airline maintenance including the identification of top human problems or failures. As per one such study (BASI, 1997), the top human failures concerning maintenance in aircraft over 5700 Kg were as follows:

- Installation of incorrect components or parts;
- Poor lubrication;
- Discrepancies in electrical wiring;
- Unsecured fuel/oil caps and refuel panels;
- Fitting of wrong parts;
- Failure to remove landing gear ground lock pins prior to departure;
- Unsecured cowlings, access panels, and fairings; and
- Loose objects left in the aircraft.

25.7 Mathematical Models for Performing Maintenance Error Analysis in Engineering Systems

Over the years many mathematical models have been developed to perform maintenance error analysis in engineering systems (Dhillon, 2002, 2006). Two such models, developed by using the Markov method, are presented below (Dhillon, 1988, 2002).

25.7.1 Model I

This model represents an engineering system that can fail due to maintenance error or malfunctioning of its parts. The system state space diagram is shown in Figure 25.5. Numerals in circle and boxes denote system states and the model is subjected to the following assumptions:

- The failed system is repaired and preventive maintenance is performed periodically;
- Maintenance error and other failure, rates are constant; and
- The failed system repair rates are constant and the repaired system is as good as new.

The following symbols are associated with the model:

i is system state, $i = 0$ means the system working normally, $i = 1$ means the system failed due to maintenance error, $i = 2$ means the system failed due to non-maintenance error (e.g., hardware failure);

$P_i(t)$ is the probability that the system is in state i at time t , for $i = 0, 1, 2$;

λ_m is the constant system maintenance error rate;

μ_m is the constant system repair rate from state 1;

λ is the constant system non-maintenance error failure rate; and

μ is the constant system repair rate from state 2.

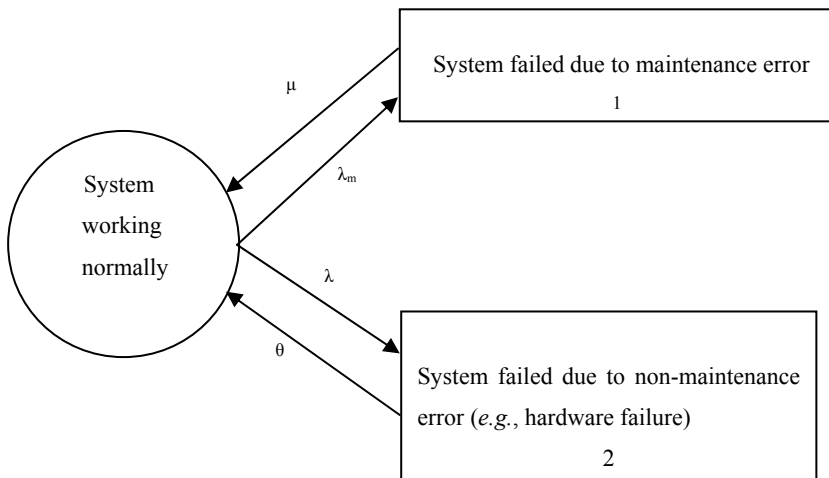


Figure 25.5. System state space diagram

With the aid of the Markov method from Figure 25.5, we get the following equations (Shooman, 1968):

$$\frac{dP_0(t)}{dt} + (\lambda_m + \lambda)P_0(t) = \mu_m P_1(t) + \theta P_2(t) \quad (25.6)$$

$$\frac{dP_1(t)}{dt} + \mu_m P_1(t) = \lambda_m P_0(t) \quad (25.7)$$

$$\frac{dP_2(t)}{dt} + \theta P_2(t) = \lambda P_0(t) \quad (25.8)$$

at time $t = 0$, $P_0(0) = 1$, $P_1(0) = 0$, and $P_2(0) = 0$.

By solving Equations 25.6 – 25.8, we obtain

$$P_0(t) = \frac{\theta \mu_m}{x_1 x_2} + \left[\frac{(x_1 + \mu_m)}{x_1 (x_1 - x_2)} \right] e^{x_1 t} - \left[\frac{(x_2 + \mu_m)(x_2 + \theta)}{x_2 (x_1 - x_2)} \right] e^{x_2 t} \quad (25.9)$$

where

$$x_1, x_2 = \frac{-B \pm [B^2 - 4(\mu_m \theta + \lambda_m \theta + \lambda \theta)]^{1/2}}{2}$$

$$B = \lambda_m + \lambda + \mu_m + \theta$$

$$x_1 x_2 = \theta \mu_m + \theta \lambda_m + \lambda \mu_m$$

$$x_1 + x_2 = -(\lambda_m + \lambda + \mu + \mu_m)$$

$$P_1(t) = \frac{\theta \lambda_m}{x_1 x_2} + \left[\frac{\lambda_m x_1 + \theta \lambda_m}{x_1 (x_1 - x_2)} \right] e^{x_1 t} - \left[\frac{(\theta + x_2) \lambda_m}{x_2 (x_1 - x_2)} \right] e^{x_2 t} \quad (25.10)$$

$$P_2(t) = \frac{\lambda \mu_m}{x_1 x_2} + \left[\frac{\lambda x_1 + \lambda \mu_m}{x_1 (x_1 - x_2)} \right] e^{x_1 t} - \left[\frac{(\mu_m + x_2) \lambda}{x_2 (x_1 - x_2)} \right] e^{x_2 t} \quad (25.11)$$

The probability of system failure due to maintenance error at time t is given by Equation 25.10.

The time dependent system availability is given by

$$AV_s(t) = P_0(t) = \frac{\theta \mu_m}{x_1 x_2} + \left[\frac{(x_1 + \mu_m)(x_1 + \theta)}{x_1 (x_1 - x_2)} \right] e^{x_1 t} - \left[\frac{(x_2 + \mu_m)(x_2 + \theta)}{x_2 (x_1 - x_2)} \right] e^{x_2 t} \quad (25.12)$$

where $AV_s(t)$ is the system availability at time t .

As t becomes very large, the system steady state availability from Equation 25.12 is given by

$$AV_s = \frac{\theta \mu_m}{x_1 x_2} \quad (25.13)$$

where AV_s is the system steady state availability.

Similarly, for very large value of t , the steady state probability of system failure due to maintenance error from Equation (25.10) is

$$P_1 = \frac{\theta \lambda_m}{x_1 x_2} \quad (25.14)$$

where P_1 is the steady state probability of system failure due to maintenance error.

For $\theta = \mu_m = 0$, from Equations 25.9 – 25.11, we obtain

$$P_0(t) = e^{-(\lambda_m + \lambda)t} \quad (25.15)$$

$$P_1(t) = \frac{\lambda_m}{\lambda + \lambda_m} \left[1 - e^{-(\lambda_m + \lambda)t} \right] \quad (25.16)$$

$$P_2(t) = \frac{\lambda}{\lambda + \lambda_m} \left[1 - e^{-(\lambda_m + \lambda)t} \right] \quad (25.17)$$

The system reliability from Equation 25.15 is

$$R_s(t) = P_0(t) = e^{-(\lambda_m + \lambda)t} \quad (25.18)$$

where $R_s(t)$ is the system reliability at time t .

The system mean time to failure (MTTFs) is expressed by (Shooman, 1968; Dhillon, 2000)

$$\begin{aligned} MTTF_s &= \int_0^{\infty} R_s(t) dt \\ &= \int_0^{\infty} e^{-(\lambda_m + \lambda)t} dt = \frac{1}{\lambda_m + \lambda} \end{aligned} \quad (25.19)$$

25.7.2 Model II

This model represents a system whose performance is degraded by the occurrence of human error, but it fails only due to non-maintenance error-related failures. The system state space diagram is shown in Figure 25.6. Numerals in circle, diamond, and box denote system states and the following assumptions are associated with the model:

- The occurrence of maintenance error can only result in system degradation, but not failure;
- The system can fail from its degraded state due to failures other than Maintenance errors;
- The totally or partially failed system is repaired and the repaired system is as good as new; and
- The system maintenance error, non-maintenance error failure, and repair rates are constant.

The following symbols are associated with the model:

i is the system state; $i = 0$ means the system working normally, $i = 1$ means the system degraded due to maintenance error, $i = 2$ means the system failed;
 $P_i(t)$ is the probability that the system is in state i at time t , for $i = 0, 1, 2$;
 λ is the system constant failure rate;
 μ is the system constant repair rate from state 2 to state 0;
 λ_m is the system constant maintenance error rate that causes system degradation;

μ_m is the system constant repair rate from the system degradation state (*i.e.*, state 1);

λ_d is the constant failure rate from the system degradation state 1 to system failed state 2; and

μ_d is the system constant repair rate from state 2 to state 1.

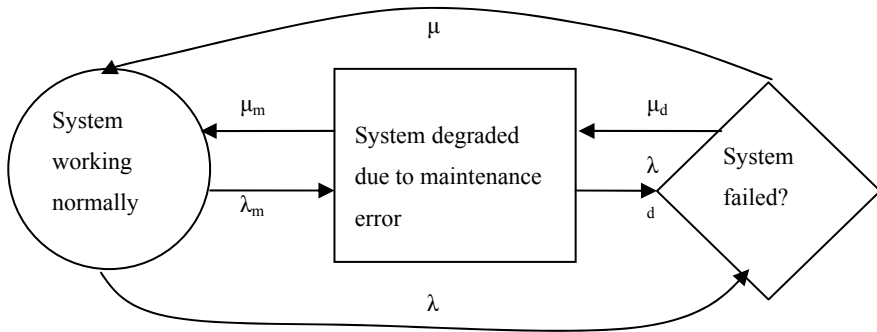


Figure 25.6. System state space diagram

By applying the Markov method, we write down the following system of equations for the diagram in Figure 25.6 (Shooman, 1968):

$$\frac{dP_0(t)}{dt} + (\lambda + \lambda_m)P_0(t) = P_1(t)\mu_m + P_2(t)\mu \quad (25.20)$$

$$\frac{dP_1(t)}{dt} + (\mu_m + \lambda_d)P_1(t) = P_0(t)\lambda_m + P_2(t)\mu_d \quad (25.21)$$

$$\frac{dP_2(t)}{dt} + (\mu + \mu_d)P_2(t) = \lambda P_0(t) + \lambda_d P_1(t) \quad (25.22)$$

at time $t = 0$, $P_0(0) = 1$, $P_1(0) = 0$, and $P_2(0) = 0$.

By solving Equations 25.20 – 25.22, we get

$$P_0(t) = \frac{(\mu\mu_m + \mu\lambda_d + \mu_d\mu_m)}{y_1 y_2} + \left[\frac{\mu_m y_1 + \mu y_1 + \mu_d y_1 + y_1 \lambda_d + y_1^2 + \mu\mu_m + \mu\lambda_d + \mu_m \mu_d}{y_1 (y_1 - y_2)} \right] e^{y_1 t} + \left\{ 1 - \left(\frac{\mu\mu_m + \mu\lambda_d + \mu_d\mu_m}{y_1 y_2} \right) - \left[\frac{y_1 \mu_m + y_1 \mu + y_1 \mu_d + y_1 \lambda_d + y_1^2 + \mu\mu_m + \mu\lambda_d + \mu_m \mu_d}{y_1 (y_1 - y_2)} \right] \right\} e^{y_2 t} \quad (25.23)$$

where

$$y_1, y_2 = \frac{-A \pm \left[A^2 - 4(\mu\mu_m + \mu\lambda_d + \mu_m \mu_d + \lambda_m \mu + \lambda_m \mu_d + \lambda_m \lambda_d + \lambda \mu_m + \lambda \mu_d + \lambda \lambda_d) \right]^{1/2}}{2}$$

$$\begin{aligned}
A &= \lambda + \lambda_m + \lambda_d + \mu + \mu_m + \mu_d \\
y_1 y_2 &= \mu \mu_m + \mu \lambda_d + \mu_m \mu_d + \lambda_m \mu + \lambda_m \mu_d + \lambda_m \lambda_d + \lambda \mu_m + \lambda \mu_d + \lambda \lambda_d \\
P_1(t) &= \frac{(\mu \lambda_m + \lambda_m \mu_d + \lambda \mu_d)}{y_1 y_2} + \left[\frac{(y_1 \lambda_m + \lambda_m \mu + \lambda_m \mu_d + \lambda \mu_d)}{y_1 (y_1 - y_2)} \right] e^{y_1 t} \\
&\quad - \left[\frac{(\lambda_m \mu + \lambda_m \mu_d + \lambda \mu_d)}{y_1 y_2} + \frac{(y_1 \lambda_m + \lambda_m \mu + \lambda_m \mu_d + \lambda \mu_d)}{y_1 (y_1 - y_2)} \right] e^{y_2 t}
\end{aligned} \tag{25.24}$$

$$\begin{aligned}
P_2(t) &= \frac{(\lambda_m \lambda_d + \mu_m \lambda + \lambda \lambda_d)}{y_1 y_2} + \left[\frac{(y_1 \lambda + \lambda_m \lambda_d + \lambda \mu_m + \lambda \lambda_d)}{y_1 (y_1 - y_2)} \right] e^{y_1 t} \\
&\quad - \left[\frac{(\lambda_m \lambda_d + \mu_m \lambda + \lambda \lambda_d)}{y_1 y_2} + \frac{(y_1 \lambda + \lambda_m \lambda_d + \mu_m \lambda + \lambda \lambda_d)}{y_1 (y_1 - y_2)} \right] e^{y_2 t}
\end{aligned} \tag{25.25}$$

The probability of system degradation at time t due to maintenance error is given by Equation 25.24. As t becomes very large, it reduces to

$$P_1 = \frac{(\lambda_m \mu + \lambda_m \mu_d + \lambda \mu_d)}{y_1 y_2} \tag{25.26}$$

where P_1 is the steady state probability of system degradation due to maintenance error.

The overall time dependent system availability, $AV_s(t)$, with maintenance error is given by

$$AV_s(t) = P_0(t) + P_1(t) \tag{25.27}$$

As t becomes very large, the system steady state availability from Equation 25.27 is given by

$$AV_s = \frac{(\mu_m \mu + \lambda_d \mu + \mu_m \mu_d + \mu \lambda_m + \lambda_m \mu_d + \lambda \mu_d)}{y_1 y_2} \tag{25.28}$$

where AV_s is the system steady state availability with maintenance error.

25.8 Useful Guidelines to Reduce the Occurrence of Human Error in Maintenance

Over the years, professionals working in the field have developed various guidelines to reduce the occurrence of human error in maintenance. This section presents guidelines developed to reduce the occurrence of human error in the area of airline maintenance. Many of these guidelines can also be used in other maintenance areas as well. The guidelines cover ten areas as shown in Figure 25.7 (BASI, 1997).

Four guidelines that cover procedures are as follows:

- Examine work practices periodically to ensure that they do not differ significantly from actual formal procedures;
- Examine documented maintenance procedures and practices periodically to ensure that they are consistent, accessible, and realistic;
- Ensure that standard work practices are followed across all areas of maintenance; and
- Evaluate the ability of checklists in regard to assisting maintenance personnel in performing routine operations such as preparing an aircraft for towing or activating hydraulics.

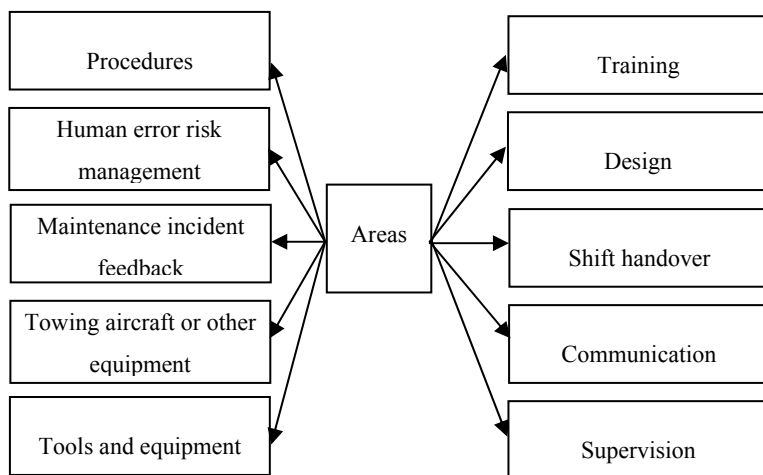


Figure 25.7. Areas covered by guidelines for reducing human error in maintenance

There are two guidelines concerning design: (1) actively seek information on errors occurring during maintenance operations to provide input in the design phase and (2) ensure that manufacturers give appropriate attention to maintenance-related human factors during the design process. Three guidelines pertaining to the area of risk management are as follows:

- Avoid performing the same maintenance task on similar redundant items or systems simultaneously;
- Review formally the adequacy of defenses designed into the system to detect maintenance errors; and
- Consider the need to disturb normally operating system to carry out non-essential periodic maintenance inspections, if there is maintenance error occurrence risk associated with a disturbance.

Two guidelines associated with training are (1) consider introducing crew resource management for maintenance personnel and others interacting with them and (2) provide appropriate refresher training to maintenance personnel with

emphasis on company procedures. One guideline pertaining to supervision is to recognize that supervision and management oversight need to be strengthened, particularly in the final hours of each shift because during this period the occurrence of errors becomes more likely. The following two guidelines are associated with tools and equipment:

- Review the system by which equipment is maintained for removing unserviceable equipment from service and repairing it rapidly; and
- Ensure that lockout devices are stored in such a way that it becomes immediately apparent, if they are left in place inadvertently.

Two guidelines pertaining to communication and towing aircraft (or other equipment) areas are (1) ensure that adequate systems are in place for disseminating important information to maintenance personnel so that changing procedures or repeated errors are considered properly and (2) review the procedures and equipment used for towing to and from maintenance facilities, respectively. One important guideline concerning shift handover is to ensure the adequacy of shift handover practices in regard to documentation and communication, so that all on going tasks are transferred correctly across all shifts.

Finally, the two guidelines covering the maintenance incident feedback area are as follows:

- Ensure that engineering training school receives feedback on recurring maintenance incidents on a regular basis, so that proper corrective measures for these problems are targeted; and
- Ensure that management receives regular and structured feedback on maintenance incidents with emphasis on the underlying conditions or latent failures that play a pivotal role in promoting such incidents.

References

- 1997 DOD Budget (1996) Potential Reductions to Operation and Maintenance Program, United States General Accounting Office, Washington, D.C.
- AMCP 706-134 (1972) Maintainability Guide for Design, United States Army Material Command, Department of the Army, Washington, D.C.
- AMCP 706-132 (1975) Engineering Design Handbook: Maintenance Engineering Techniques. Department of Army, Washington, D.C.
- Anderson, RT, Neri L (1990), Reliability Centered Maintenance: Management and Engineering Methods, Elsevier Applied Sciences, London.
- BASI (1997) Human Factors in Airline Maintenance: A Study of Incident Reports, Bureau of Air Safety Investigation (BASI), Department of Transport and Regional Development, Canberra, Australia.
- Beech HR, Burns LE, Sheffield BF (1982) A Behavioral Approach to the Management of Stress. John Wiley and Sons, New York.
- Blanchard BS (1992), Logistics Engineering and Management, Prentice-Hall, Englewood Cliffs, NJ.

- Christensen JM and Howard JM (1981) Field Experience in Maintenance. In *Human Detection and Diagnosis of System Failures*, edited by J. Rasmussen and W.B. Rouse, Plenum Press, New York, pp 111–113.
- Circular 243 – AN/151 (1995) *Human Factors in Aircraft Maintenance and Inspection*, International Civil Aviation Organization, Montreal, Canada.
- Dhillon BS (1986) *Human Reliability: With Human Factors*. Pergamon Press, New York.
- Dhillon BS (1988) *Mechanical Reliability: Theory, Models, and Applications*. American Institute of Aeronautics and Astronautics, Washington, D.C.
- Dhillon BS (2002) *Engineering Maintenance: A Modern Approach*. CRC Press, Boca Raton, Florida.
- Dhillon BS (2006) *Maintainability, Maintenance, and Reliability for Engineers*, CRC Press, Boca Raton, Florida.
- Dhillon BS and Liu Y (2006) Human Error in Maintenance: A Review. *J of Qual in Maint Eng* 12(1): 21–36.
- Hagen EW Editor (1976) *Human Reliability Analysis*. Nucl Saf 17: 315–326.
- Hammer W (1980) *Product Safety Management and Engineering*, Prentice Hall, Englewood Cliffs, New Jersey.
- Latino CJ (1999) *Hidden Treasure: Eliminating Chronic Failures Can Cut Maintenance Costs up to 60%*, Report, Reliability Center, Hopewell, Virginia.
- McKenna T and Oliverson R (1997) *Glossary of Reliability and Maintenance Terms*. Gulf Publishing Company, Houston, Texas.
- Meister D (1962) The Problem of Human-Initiated Failures, *Proceedings of the 8th National Symposium on Reliability and quality Control*, pp. 234–239.
- Meister D (1976) *Human Factors: Theory and Practice*, John Wiley and Sons, New York.
- MIL-STD-721B (1966) *Definitions of Effectiveness Terms for Reliability, Maintainability, Human Factors, and Safety*, Department of Defense, Washington, D.C., August.
- Naresky JJ (1970) Reliability Definitions. *IEEE Transactions on Reliability*, 19: 198–200.
- Omdahl TP (1988) *Reliability, Availability, and Maintainability (RAM) Dictionary*. ASQC Quality Press, Milwaukee, Wisconsin.
- Regulinski TL and Askren WB (1969) *Mathematical Modeling of Human Performance Reliability*, *Proceedings of the Annual Symposium on Reliability*, pp 5–11.
- Robinson JE, Deutsch WE, Rogers JG (1970) The Field Maintenance Interface Between Human Engineering and Maintainability Engineering. *Hum Factors* 12: 253–259.
- Russell PD (1994) *Management Strategies for Accident Prevention*. *Air Asia* 6: 31–41.
- Sauer D, Campbell WB, Potter NR, Askern WB (1976) *Relationships Between Human Resource Factors and Performance on Nuclear Missile Handling Tasks*, Report No. AFHRL-TR-76-85/AFWL-TR-76-301, Air Force Human Resources Laboratory/Air Force Weapons Laboratory, Wright-Patterson Air Force Base, Ohio.
- Shooman ML (1968) *Probabilistic Reliability: An Engineering Approach*, McGraw Hill Book Company, New York.
- The Volume Library (1993) *A Modern Authoritative Reference for Home and School Use*, The South-Western Company, Nashville, Tennessee.
- Tripp EG (1999) *Human Factors in Maintenance*. B/CA, July 1999, pp 1–10.

Human Error in Maintenance – A Design Perspective

Clive Nicholas

26.1 Introduction

Emphasis on the elimination or reduction of human error in maintenance and its consequences is a relatively recent phenomenon. Human error in maintenance can result in maintenance error that may potentially degrade the performance of technical systems and possibly give rise to extremely serious safety and economic consequences.

Much of the initial focus in addressing human error has been placed upon the role of the system operator through personnel training, through the adoption of procedures and practices and through regulation. More recently, there has been a growing awareness of the impact that system design can have on human error in maintenance. This chapter examines how potential human error in maintenance can be systematically analyzed to develop specific design strategies that can be used to reduce the occurrence of human error in maintenance and to mitigate its consequences. The content of the chapter is based upon the author's extensive experience of developing and applying such analysis and design strategies in the aerospace industry where the principles and methodology discussed have been employed in the design of civil and military aircraft. However, the principles and methodology are generic and can be applied to other technical systems where the potential for human error is present in maintenance activities.

Mechanics and engineers who are exposed to airplanes in their daily work must constantly look for what is wrong with the design, what is wrong or out of place on the in-service airplane, what is wrong with maintenance data, procedures, processes. What is broken, leaking, corroded, deformed, chaffing, cracked, etc.? Always assume that something is wrong that will compromise safety.

Jack Hessburg – Former Chief Mechanic,
New Airplanes Boeing Commercial Aircraft Group (Hessburg, 2001)

The aircraft maintenance process consists of a flow of tasks designed to maintain the safe and economic operation of the aircraft. Maintenance tasks

typically include removal, installation, servicing, rigging, inspection and other scheduled maintenance.

The execution of any maintenance task involves the possibility of human error. Human error in aircraft maintenance is the consequence of a complex interaction of many factors including system and maintenance task design, maintenance personnel and other resources, maintenance organisation, and the physical environment in which the maintenance occurs.

Clearly aircraft designers cannot eliminate every potential cause of human error in the maintenance process of the operator. However, it is possible to have a significant impact upon the possibility of human error through the design of aircraft systems or items (*i.e.*, the maintainability characteristics of the aircraft) and the design of the maintenance process (*i.e.*, the actual maintenance tasks and supporting resources and activities).

26.2 Human Error in Aircraft Maintenance

Human error in aircraft maintenance is the unintentional act of performing a maintenance task incorrectly that can potentially degrade the performance of the aircraft. The physical effect of failure to perform a task correctly is a maintenance error. For example, if a maintainer, working in limited conditions, fails to complete the task correctly due to personal limitations of physical strength to lift and place a component correctly, the resulting maintenance error could be an incorrect installation leading to potential failure of the component.

Most human errors in aircraft maintenance are the result of unintentional inappropriate actions that lead to maintenance error in a particular set of circumstances. There are also intentional erroneous actions on the part of the maintainer when, for some reason, it is either considered to be the correct action or a better way of performing a maintenance task. In each case the maintainer is acting in a rational manner and with good intent to perform safe and effective maintenance.

It should also be recognised that human error does not necessarily always result in degradation of the aircraft. An error can often be detected and recovered before it results in consequential degradation. Error detection is significant because quite clearly it is important that the error is detected when the aircraft is in maintenance rather than when it is in service.

Human behavior is variable and is determined by a considerable range of factors that can vary significantly in different conditions and environments. Common factors can produce different responses and effects. Individual behaviors do not display uniformity and the designer would find it difficult to generate a design solution that would be applicable to the individual behaviors of maintainers. However, when designing an aircraft system or component the designer can address common patterns of behavior manifest in reasonably foreseeable maintenance errors.

26.3 Significance of Maintenance Error

The significance of human error in maintenance is dependent upon its safety and economic consequences. In aircraft operations, safety is of paramount importance. The safety and economic effects of maintenance error can range from little or no consequence, through physical damage of equipment and personal injury, to catastrophic loss of the aircraft and loss of life.

The safety impact of human error in maintenance is illustrated by the following examples.

Example 26.1: Maintenance failure probable cause of crash (NTSB, 2002)

On February 16, 2000, Emery Worldwide Airlines, Inc., (Emery) flight 17, a McDonnell Douglas DC-8-71F (DC-8), N8079U, crashed in an automobile salvage yard shortly after takeoff, while attempting to return to Sacramento Mather Airport (MHR), Rancho Cordova, California, for an emergency landing.

The flight departed MHR with two pilots and a flight engineer on board. The three flight crewmembers were killed, and the airplane was destroyed.

The National Transportation Safety Board (NTSB) determined that the probable cause of the accident was a loss of pitch control resulting from the disconnection of the right elevator control tab. The disconnection was caused by the failure to secure and inspect the attachment bolt properly.

The NTSB concluded that the bolt attaching the accident airplane's right elevator control tab was improperly secured and inspected, either during the most recent "D" inspection or during subsequent maintenance.

As a result of the accident, the NTSB issued 15 safety recommendations to the Federal Aviation Authority (FAA), including a recommendation for redesign of elevator control tab installations and retrofit.

(Adapted from NTSB Aircraft Accident Report NTSB/AAR-03/02, 2002)

Example 26.2: Maintenance error blamed in fighter pilot's death (CBC, 2005)

A maintenance error led to mechanical failure in a crash that led to the death of a Canadian fighter pilot in South Carolina on June 28 2004. The F-18 Hornet crashed while landing at a US Marine Corps base after a 10-h flight from Denmark. A joint US-Canadian investigation concluded that a problem with the landing gear caused the aircraft to skid off the runway and flip over. The investigation report identified an improperly installed landing gear strut as the cause of the accident. The work on the landing gear was carried out more than a month before the crash.

(Adapted from CBC Canada News Updated 11 March 2005)

Example 26.3: Improper maintenance caused F-16D crash (F-16.net, 2005)

A maintenance crew's failure to put seals on an engine part caused an F-16D to crash into a Charleston marsh in the US on April 18 2005. The pilot and passenger ejected and sustained minor injuries. The aircraft was destroyed on impact.

The Accident Investigation Board Report indicated that the high-pressure turbine rotor failed, resulting in significant loss of thrust. The pilot attempted three

engine restarts while maneuvering for a straight-in flameout runway approach. Unable to reach the runway safely, the pilot steered toward an unpopulated marshland and initiated a dual ejection.

The report stated that the seals should have been installed about a year before the crash. Without the seals, excess heat built up in the engine, causing the metal to become brittle.

(Adapted from F-16.net 23 August 2005)

The economic significance of human error in maintenance is often overlooked or included in other costs that do not quantify the consequential cost of error. Examples of specific costs are illustrated in the estimates for UK civil aircraft operators shown in Table 26.1.

Table 26.1. Cost estimates per operator (subject to use of multipliers to reflect aircraft types operated) per annum per aircraft (Royal Aeronautical Society, 2000)

Airborne turnback	£10,000
Doing the same job twice	£12,500
Wrong parts installed	£20,000
Cross-connected systems	£20,000
Mis-diagnosed defects	£40,000
Ramp damage	£40,000

Discussions with maintenance engineers, mechanics, and technicians have shown that there is considerable anecdotal evidence of the occurrence of human error leading to maintenance error. Formal studies and surveys of aircraft accidents and fatalities have provided documentary evidence to support this.

In a detailed analysis of 93 major worldwide accidents that occurred between 1959 and 1983 (Sears, 1993), maintenance and inspection were factors in 12% of accidents as shown in Table 26.2.

It has been estimated (Marx 1998) that in the U.S.A. the number of aircraft per year dispatched into revenue service in a technically unairworthy condition because of maintenance error is approximately 48,800. Considered on a per aircraft basis, the average airplane would see roughly seven airworthiness-related maintenance errors per year.

Table 26.2. Analysis of 93 major worldwide accidents 1959–1983 (Sears, 1993)

Causes/major contributory factors	Percentage of accidents in which this was a factor
Pilot deviated from standard procedures	33
Inadequate cross-check by second crew member	26
Design faults	13
Maintenance and inspection deficiencies	12
Absence of approach guidance	10
Captain ignored crew inputs	10
Air traffic control failures or errors	9
Improper crew response during abnormal conditions	9
Insufficient or incorrect weather information	8
Runways hazards	7
Improper decision to land	6
Air traffic control/crew communication deficiencies	6

The statistic of 48,800 aircraft dispatched in an unairworthy condition does not necessarily imply that there were 48,000 unsafe aircraft dispatched each year but rather that they were dispatched out of conformity with their type design because of error on the part of the maintainer – the aircraft were in a condition that was not intended by the maintainer.

Of 14 US National Transportation Safety Board investigations of large aircraft accidents (Goglia, 2000), 7 had maintenance as a major contributory factor (*i.e.*, 50%). The study suggested that either maintenance problems are on the increase or that, as improvements are made in aircraft design, pilot training, Air Traffic Control, *etc.*, the proportion of accidents attributable to these factors is lower and the proportion attributable to poor maintenance is consequently higher.

Evidence based on UK Civil Aviation Authority Mandatory Occurrence Reporting Scheme statistics (Courteney, 2001) has indicated a continuing rise in the number of reportable maintenance errors per million flights. The trend is shown in Figure 26.1. In addition, it was observed that such a rise would be compounded by increasing traffic to make absolute numbers of errors show an accelerating trend.

Data produced by the Boeing Commercial Airplane Group (Boeing Commercial Airplane Group, 2003) suggests that maintenance accounted for 3% of accidents by primary cause represented as Hull Losses in the Worldwide Commercial Jet Fleet 1993–2002. The results of the study are shown in Figure 26.2.

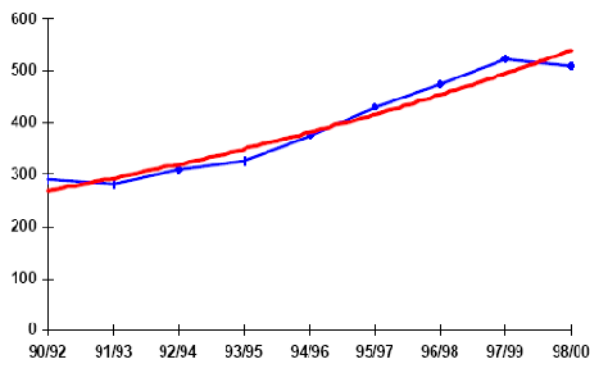


Figure 26.1. Maintenance error MORs to UK registered public transport aeroplanes > 5700kg mtwa per million flights (shown as 3 year moving average)

Accidents by Primary Cause*
Hull Loss - Worldwide Commercial Jet Fleet - 1993 through 2002

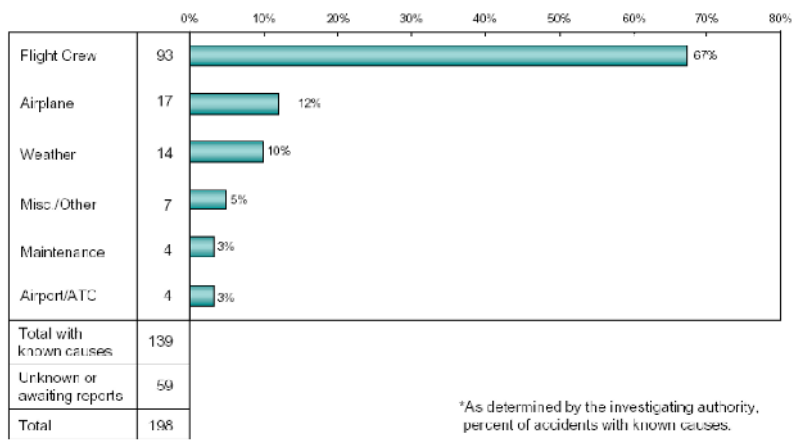


Figure 26.2. Accidents by primary cause, hull loss in the worldwide commercial jet fleet, 1993–2002 (Boeing Commercial Airplane Group, 2003)

The Flight International Safety Review of 2003 suggested that in that year, technical and maintenance faults took over from controlled flight into terrain (CFIT) as the biggest cause of fatal airliner accidents (Figure 26.3).

These studies and surveys provide evidence of the occurrence of maintenance error but it is important not to misinterpret the precise nature of the data. Some of the data indicate that maintenance error, as either a cause or contributory factor, is a relatively small percentage, whilst other data suggest a greater significance and, indeed, an increasing significance.

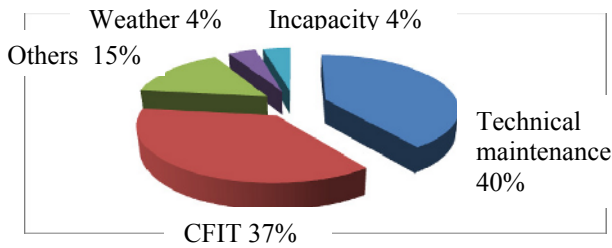


Figure 26.3. Fatal accidents by category 2003 (Flight International, 2003)

Whatever the results of the studies, it is a fact that the data do not reflect aviation as a whole. Furthermore, within those sectors that are covered it is possible that maintenance error is inadequately recorded or is subsumed in other accident causes.

There are two important conclusions to draw. First, maintenance error can and does occur and therefore there is a probability that it will occur in the future. Second, although the precise magnitude and frequency of occurrence maintenance error over time remain uncertain, general behavioral patterns exist in maintenance that will assist in systematically addressing the causes and consequences of maintenance error as an integral part of the aircraft design process.

26.4 Design Impact

The safe and economic completion of an aircraft maintenance task depends upon the interaction and inter-relationships of the design characteristics of the aircraft and its operation in a specific environment. Design characteristics of the aircraft include technical systems, components and items. They also include the consequent design of maintenance tasks, procedures, manuals, tools, equipment and initial training of maintainers. Operation will include the characteristics of maintenance personnel, the maintenance organisation and the physical environment within which they work

The human and the aircraft interact through the maintenance task. The purpose of the aircraft is to provide a set of functions that enable its operation to deliver a safe flight that departs and arrives on schedule. The aircraft's ability to deliver safe flights is sustained through maintenance to ensure that it functions as and when required. The organization and resources for operation and maintenance are provided through support organizations.

The operation, maintenance and support of an aircraft are made up of related processes, which consist of tasks carried out by humans using available resources.

A maintenance task can be described in the following terms:

- A maintenance task is any specified set of maintenance actions that is performed to maintain the required function of an aircraft component or system;

- The set of maintenance actions is related by their task requirement and their sequential occurrence in time;
- The execution of maintenance tasks involves human actions that comprise of some combination of cognitive (“thinking”) and physical action (“doing”); and
- Each task requires an expected level of maintenance performance to be complete each action and the task as a whole.

The successful completion of a maintenance task as specified therefore involves:

- The human performance and limitations (*e.g.*, vision, hearing, physique, perception, memory, fatigue, *etc.*);
- System and process design: the demands placed on human performance that are the result of design (*e.g.*, operation, maintenance and support task and resource demands);
- System and process operation: the demands placed on human performance that are a result of operation (*e.g.*, organisation, procedures, *etc.*); and
- Physical environment: the demands placed on human performance that are a result of the physical environment in which the task is performed (*e.g.*, climate, temperature, noise, illumination, *etc.*).

It is clear that aircraft designers are not in a position to control all these factors. However, they can have an impact upon them through design solutions that influence the potential for human error in maintenance. This can be achieved by developing an understanding of the types of maintenance tasks that are a consequence of aircraft design; the maintenance errors that can emanate from these tasks; the forms of human error that can give rise to maintenance error; and human performance influencing factors and actively integrating into design specific solutions that address the potential for human error in maintenance as a consequence of design.

26.5 Analysis Required for Design Solutions

Analyzing design for potential human error in maintenance and developing strategies to deal with error and its consequences requires a systematic approach. The logic flow for such an approach is shown in Figure 26.4 and is outlined below.

This logic flow can be transposed into a practical analytical tool by applying the general principles of Failure Modes and Effects Analysis (FMEA). This method has traditionally been used for the analysis of technical failures but can be adapted to provide a structure for the qualitative analysis of human error in maintenance and a means of recommending strategies to eliminate or mitigate maintenance errors and their effects. It can be described as a form of Process FMEA that analyzes the maintenance process for potential errors.

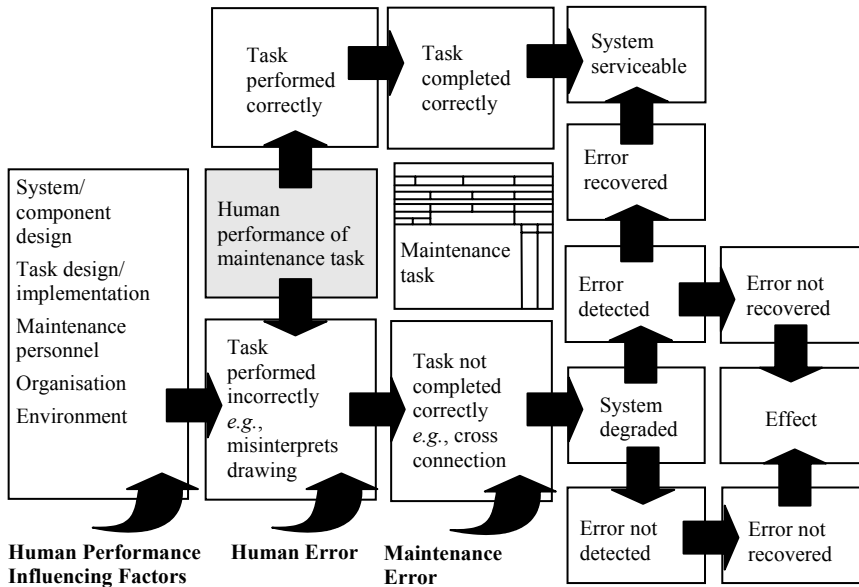


Figure 26.4. Logic flow for analysis of human error in aircraft maintenance

The general steps involved in the analysis can be summarised as follows:

- Describe the candidate item (selected using specified criteria - *e.g.*, safety critical);
- Describe the item failure condition;
- Identify maintenance tasks that are carried out on the item;
- Identify possible maintenance error;
- Determine the primary cause;
- Determine whether there is an immediate consequence of error that is evident to the maintainer;
- Determine whether error could lead to the failure condition;
- Determine whether a functional or operational test is required that could Detect maintenance error;
- Determine whether other indication of maintenance error is apparent to the maintainer;
- Determine whether there is an indication of maintenance error to the operator before operation;
- Determine whether there is a subsequent indication of maintenance error to the operator; and
- Develop solutions.

The analysis can be conducted using a simple worksheet as illustrated in Figure 26.5. Alternatively, the process can be mechanised using appropriate computer software.

Ref	System Component		Error			Consequences		Detection				Solutions	Other
Ref No.	System/ component Description	Failure Description	Maintenance tasks	Maintenance Error	Cause	Immediate evidence of Error	Effect of Error (can the error create the failure?)	Functional or operational test required	Other indication to maintainer	Indication to operator before operation	Subsequent indication to operator	Recommendations	Comments

Figure 26.5. Example of analysis worksheet

26.5.1 Maintenance Tasks

Aircraft maintenance involves the inspection, overhaul, repair, preservation and replacement of parts to maintain the functioning of the aircraft and to ensure that all safety and regulatory requirements are met.

Examples of aircraft maintenance tasks are shown in Table 26.3.

Table 26.3. Aircraft maintenance tasks

Category	Definition
Servicing	Replenishment of consumable fluids, cleaning, washing, painting
Lubrication	The act of installing or replenishment of lubricants
Inspection	Examination of an item against a defined physical standard
General visual	An inspection that will detect obvious unsatisfactory conditions. The inspection may require the removal of fairings, fillets, access doors or panels. Workstands, ladders, <i>etc.</i> , may be required to gain access
Detailed	An intensive visual examination of a specified component, assembly or system. It searches for evidence of any irregularity. Inspection aids such as mirrors, special lighting, hand lens, <i>etc.</i> , are normally employed. Surface cleaning may be required. Elaborate access procedures may be required
Special detailed	An intense examination of a specific area using special inspection equipment such as radiographic techniques, dye penetrant, eddy current, high power magnification or other NDT techniques. Elaborate access and detailed disassembly may be involved
Check	A qualitative or quantitative assessment of function
Functional	A quantitative assessment of one or more functions of an item to determine if it performs within acceptable limits
Operational	A qualitative assessment to determine if an item is fulfilling its intended function. It does not require quantitative tolerances
Visual	A failure finding observation of an item to determine if it is fulfilling its' intended function
Restoration	That work necessary to return an item to a specific standard. This may involve cleaning, repair, replacement or overhaul
Discard	Removal of an item from service

Each task is defined in the form of a specific procedure that involves a sequence of maintenance actions that are originally identified and documented through the process of Maintenance Task Analysis together with personnel, resource and time factors. The precise form of the maintenance task will determine the type of potential maintenance errors that might occur as a consequence of incorrect completion.

26.5.2 Maintenance Errors

Aircraft maintenance can be a very complex process that places considerable demands upon the human to perform at the level required by the maintenance task. Given the fact that maintenance also often occurs in an environment that is essentially hostile to the human maintainer, the conditions for maintenance error abound.

Although there can potentially be many forms that a maintenance error can take, empirical evidence indicates that there are common patterns of error. Frequently occurring maintenance errors include:

- Wrong part installed;
- Fault not found by inspection;
- Incomplete installation;
- Cross connection;
- Fault not detected;
- Wrong orientation;
- Access not closed;
- Wrong fluid;
- Servicing not performed;
- Fault not found by test;
- System not deactivated; and
- Material left in aircraft.

Evidence of the occurrence of these forms of maintenance error is available from various sources. The UK Civil Aviation Authority, for example, issued a list in 1992 of frequently recurring maintenance discrepancies based on Mandatory Occurrence Reports. The problems identified (in order of frequency of occurrence) were as shown in Table 26.4.

Table 26.4. UK civil aviation authority maintenance mandatory occurrence reports Analysis 1992.

• Incorrect installation of components
• Fitting of wrong parts
• Electrical wiring discrepancies (including cross-connections)
• Loose objects (tools, <i>etc.</i>) left in aircraft
• Inadequate lubrication
• Cowling, access panels and fairings not secured
• Landing gear ground lock pins not removed before departure

An analysis by the Boeing Commercial Airplane Group of 122 documented occurrences in the period 1989–1991 involving human factors errors with likely engineering relevance found that the main categories were omissions and incorrect installation as shown in Figure 26.6.

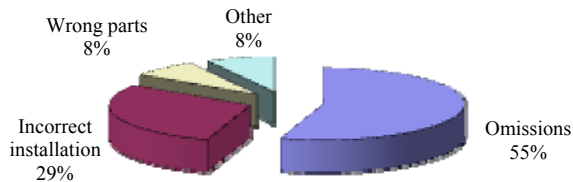


Figure 26.6. Analysis of maintenance error (Graeber and Marx, 1993)

A further breakdown of the 1993 Boeing Study figures (Reason, 1997) shows the following analysis of maintenance errors:

- Fastenings undone/incomplete (22%);
- Items left locked/pins not removed (13%);
- Caps loose or missing (11%);
- Items left loose or disconnected (10%);
- Items missing (10%);
- Tools/spare fastenings not removed (10%);
- Lack of lubrication (7%); and
- Panels left off (3%).

Based on Maintenance Error Management System (MEMS) 2002 data from several UK maintenance organisations using Maintenance Error Decision Aid (MEDA) terminology, the three top items for the categories are shown in Table 26.5.

Table 26.5. 2003 CHIRP-MES data (MEMS-MEDA, 2003)

Installation error	Fault isolation/test/inspection error	Servicing Error
3 top items		
Incomplete installation (181)	System not re/de-activated (60)	Service not performed (55)
Wrong orientation (111)	Not properly tested (58)	System not re/de-activated (24)
System not re/de-activated (87)	Not properly inspected (33)	Insufficient fluid (11)

26.5.3 Causal Factors

The potential maintenance error identified is the physical consequence of human performance of maintenance and more specifically the consequence of failure to meet the expected level of task performance. Therefore the fundamental causes of error should be examined. These are primarily driven by the demands of the maintenance task in relation to the expected human performance. The demands of the maintenance are determined by causal factors that influence human performance of the maintenance task. These factors are discussed below.

26.5.3.1 Inadequate System/Component Design

The design of the system or component will influence the form, frequency and duration of maintenance tasks carried out in operation. The complexity of design configuration, physical form, weight, location, method of installation, visual information through gauges and dials, and similar factors will play a key role in determining the demands placed upon the level of human performance required to successfully complete a maintenance task. Limitations or inadequacy in the design of systems or components may cause degradation in human performance leading to human error and consequently maintenance error.

26.5.3.2 Inadequate Task Design

Maintenance tasks differ in the physical and cognitive effort necessary for successful completion. Inadequate task design fails to take into account the demands placed on the maintainer.

The following task characteristics will impact upon the potential for human error:

- Steps;
- Sequence;
- Duration;
- Frequency;
- Personnel;
- Information;
- Documentation;
- Tools and equipment;
- Materials; and
- Environment.

Each of these task characteristics place different demands upon human capabilities and, as a consequence, can cause error when the requirements of the task exceed the capabilities and limitations of human performance.

26.5.3.3 Maintenance Personnel Performance Limitations

A primary cause of human error in maintenance is the state and condition of the human undertaking the maintenance task. Maintenance personnel limitations relate to the individual and to the abilities and attitudes that influence performance. These

attributes are determined by the physical and mental state and will show considerable variability not only between individuals but also over time.

The determinants of performance will include individual factors such as knowledge, skills, experience, physique, strength, sensory acuity, and health, and external influences that affect the individual's performance such as fatigue, stress, peer pressure and time constraints.

Error is the consequence of both the physical and cognitive capabilities and limitations of the maintainer in relation to the demands of the task. For example, failure to install a component correctly in the right alignment may be the consequence of limited physical or visual access to the location of the maintenance activity.

From a cognitive perspective, when performing a maintenance task the maintainer receives a flow of information from various sources:

- Maintainers knowledge;
- Maintenance organisation;
- Documentation;
- Aircraft information and state; and
- Maintenance environment.

The information is received by the maintainer and based on the internal processing of this information the maintainer makes decisions that result in a maintenance action (*e.g.*, to refill a reservoir).

Processing information and making decisions that lead to actions are complex cognitive processes that draw upon the maintainer's knowledge, skills, experience and perception of the situation. Usually, the more complex the flow of information, the more complex decision-making becomes.

In addition, there is generally a need to balance information against operational priorities and maintenance capabilities. Decisions will involve a trade-off amongst these and other factors.

Given this complexity of information flow and decision-making there is clearly a risk that actions taken by the maintainer may on occasion be incorrect. As a result, the maintenance task, or elements of the task, may not be completed as required and this can result in an unintended aircraft discrepancy.

Common maintainer behaviors that can result in error include:

- Memory lapses to complete actions (*e.g.*, forgetting to replace an oil cap, close an access panel or remove a tool;
- Not referring to approved maintenance documentation, abbreviating procedures (workarounds), or using informal sources of information (black books); and
- Lack of system knowledge and failures in problem-solving.

However, it is important to emphasise that not all incorrect actions will necessarily result in an aircraft discrepancy – error is an integral part of human behavior. Many incorrect actions can be recovered or corrected.

26.5.3.4 Poor Organisational Conditions

Poor organisational conditions can have a significant causal effect on the occurrence of human error in maintenance. Organisational ethos, policies, procedures and practices, management, supervision, communication, technical support and similar factors can affect the expected and actual level of performance of the maintainer.

26.5.3.5 Poor Environmental Conditions

The maintenance task is carried out in a physical environment that will have an impact upon human performance. Poor environmental conditions can be a cause of human error in maintenance. The effect on performance is both physiological and psychological. For example, maintenance might occur in work conditions that are too hot or too cold; in poor weather conditions that might include high humidity, wind, snow, and rain; or in facilities where there are high noise levels, dirt, poor lighting and ventilation.

26.6 Design Strategies and Principles

Having identified possible causes of human error in maintenance, it is clear that aircraft manufacturers can have a significant impact on the incidence of human error in maintenance through system design and, more specifically, the interface between human and machine. Aircraft can be designed to be robust to human error and the consequences of human error not only during operation but also during maintenance.

Based on the analysis of error leading to the identification of the fundamental human performance influencing factors, it is evident that the aircraft designer can have an impact on the potential for human error in maintenance through the design of elements such as:

- System;
- Component;
- Procedures and tasks;
- Tools and equipment;
- Information and documentation; and
- Initial training.

There has long been a philosophy in aircraft design that errors by maintainers are not the concern of the designer – maintainers should be trained not to make errors. That philosophy is rapidly changing. Designers have an important role to play because design characteristics have a significant impact on the form, frequency and duration of the maintenance task and have important implications for the possible occurrence of maintenance error.

As previously stated, the maintainer and the aircraft interact through the maintenance task. It is through the maintenance task that the aircraft affects the

performance of the maintainer and the maintainer affects the performance of the aircraft. The design of the system or component will influence the type, frequency and duration of maintenance tasks carried out in operation.

Key questions for the designer to consider are:

- What types of maintenance tasks does the design generate and what actions do they involve?
- How often is the maintenance task needed and how long will it take?
- What demands does the design place upon the capabilities of the maintainer to complete maintenance task? and
- Can the demands of the task exceed the possible limitations of the maintainer?

The complexity of design configuration, physical form, weight, location, access, method of installation, visual information and similar factors play an important part in determining the demands placed upon the level of maintenance performance required to complete a maintenance task successfully. Different designs will have different effects on maintenance performance. For example, the use of fewer parts may influence how easy it is to do the task – improving maintenance performance and reducing the likelihood of maintenance error.

Aircraft maintenance often involves complex processes that place considerable demands upon the maintainer to perform at the level required by the maintenance task. Maintenance often occurs in environments that also often place considerable demands upon the maintainer.

It is important to recognise the human capabilities and limitations of the maintainer and the capabilities and limitations that are inherent in any aircraft design. It involves the design of aircraft so that the relationship between the aircraft design and the maintainer effected through the maintenance task will result in optimal maintenance performance that minimises demands on maintainers that could lead to maintenance error.

The design of aircraft systems and components and the operational environment in which that design functions will influence the behavior of the maintainer – for example, how easy it is to complete the task. Design characteristics can generate tasks that are within the capabilities and limitations of the maintainer that have a potentially positive effect on maintenance performance. Equally, design characteristics can challenge the capabilities and limitations of the maintainer and have a potentially negative effect on maintenance performance. Amongst other consequences, such as decreased maintenance efficiency, this could lead to error or personal injury during maintenance.

Design can therefore affect the vulnerability of an aircraft to maintenance error and the consequences of that error. By actively integrating general principles that address maintenance error into the design process, it is possible to create design characteristics that can possibly prevent or reduce maintenance error (*e.g.*, sealed units or colour coding) or eliminate or mitigate the consequences of maintenance error (*e.g.*, isolation or partial operation).

In developing design strategies and principles that enable the practical realization of these strategies through physical design characteristics, it is

important to recognize that error is an integral and important part of fundamental human behavior – it is part of the normal cognitive and learning processes of the human. Indeed, error in itself is not inherently problematic. It is only problematic when its consequences bring about unwanted or negative consequences. Design strategies should therefore attempt to avoid errors or to contain the consequences before they become negative. Error in maintenance is a normal part of maintenance operations that can be addressed during the design process.

Design strategies may revolve around two basic approaches. The first is avoidance of error. Here the error may be completely avoided by prevention. Examples of this type of strategy include designing out operation significant maintenance tasks, the design of components that are physically impossible to assemble or install incorrectly and the use of staggered part positions that require a specific configuration or sealed units that do not require intervention.

It is also possible to reduce the frequency of occurrence of error. Examples of error frequency reduction include the use of different part numbers, colour coding, shaped switch tops, locking switches, standard display formats, standard direction of operation, convenient access panels, reduction of servicing frequency, protection against accidental damage, or lubrication points that do not require disassembly.

The second is tolerance of error. Here mechanisms to detect error, to reduce the impact of error, and to recover error may be employed. Mechanisms to detect error may include built-in tests, functional tests, illuminated test points, functionally grouped tests or warning lights. Detection error can also include initial training of the maintainer for system state recognition.

Reduction of the impact of error can be achieved through strategies such as isolation of the consequences of error, the ability for partial operation or the use of redundancy in systems or components. Recovery of error may be achieved through self-correction, the development of recovery procedures or specific training for error recovery.

Specific design objectives can be summarised as follows:

- Design that absolutely eliminates any possibility of an identified maintenance error or eliminates its consequences;
- Design that reduces the size of an identified maintenance error or reduces the extent of its consequences;
- Design that reduces how often an identified maintenance error, or how often its consequences, are likely to occur;
- Design that ensures that the maintenance error or its consequences is evident under all maintenance conditions, easy and rapid to detect, and is detected before flight; and
- Design that ensures that following a maintenance error the means to return a system to its correct state are evident, easy, and timely.

In practice, the strategies of avoidance and tolerance are complementary and it may be felt necessary to design using a combination. An error tolerant design may be combined with error avoidance mechanisms to produce a robust design. Total avoidance of error may be considered to be an ideal given the nature and variability

of human performance – error tolerance will capture and contain errors that fail avoidance mechanisms.

The general design principles discussed below provide practical means by which these strategies can be realised.

26.6.1 Appreciate the Maintainer's Perspective of the Aircraft

Designers design systems or components to deliver their required functionality. Maintainers are responsible for maintaining that functionality over the life of the aircraft whilst ensuring safety standards and operational requirements are met.

As a consequence, maintainers have a very specific perspective of an aircraft that will focus on the efficiency and safety of maintenance. Maintainers look for 'maintainer friendly aircraft' whose design characteristics enable them to achieve good maintenance performance that delivers the aircraft back into service when required by the operator and that will complete the flight in safety.

From the maintainer's perspective therefore questions arise such as:

- How long will the task take?
- Is the task complicated?
- How often is the task required?
- Do I need special training?
- Do I need special tools or equipment?
- Could I make an error?
- How will I know if things go wrong?
- Where is the item located on the aircraft?
- Is there enough space to work in?
- Can I see the item?
- Can I reach the item? and
- Where will I carry out the maintenance?

26.6.2 Design for the Aircraft Maintenance Environment

To appreciate fully the impact of design on maintenance performance it is important to understand the environment in which aircraft maintainers work. Aircraft maintenance generally takes place under conditions that are complex and very demanding.

Line maintenance, for example, is generally performed outside the hanger working on the airport ramp or apron area in all types of weather and climate, often at night with limited visibility. The environment is extremely busy with aircraft loading and servicing vehicles moving around. There is considerable noise and there are fumes from aircraft engines and APUs (auxiliary power units) running. Above all there is constant pressure to complete maintenance activities as quickly as possible to turn the aircraft around on time for departure. Operators are in the business of transporting passengers. Aircraft on the ground cost money and lose revenue for the operator.

Similarly, base maintenance that is generally carried out in the hanger involves an environment where there is a considerable amount of activity and pressure to get the job done. Having to meet exacting work schedules while still observing standard procedures and safety standards can be stressful. The hangar is generally noisy from the use of power tools and there are many fluids and substances (hydraulic fluids, cleaning compounds, fuel, paints, *etc.*) that are potentially dangerous.

Maintenance is often carried out at night when the aircraft are not in use. This means the work requires regular shift working. Requirements for overtime working and call-outs are common. Maintenance tasks can be physically demanding, involving lifting, working in uncomfortable positions or working at height on scaffolds or cherry pickers (lifts).

The aircraft maintenance environment places considerable demands upon the maintainer and upon maintenance performance. The physical environment has an impact on maintenance performance through:

- Lighting;
- Climate (dry or humid climates);
- Temperature (hot or cold temperatures);
- Weather (rain, wind, ice, snow, *etc.*);
- Fumes and toxic substances;
- Noise;
- Motion; and
- Vibration.

Clearly designers cannot directly influence the many factors present in the working environment that will affect maintenance performance. However, they can have an impact on maintenance performance by taking them into consideration during the design process and reflecting this in design solutions. For example, where maintenance tasks are carried out in extremely low temperatures it is important to consider whether a maintenance task generated by a particular design could be carried out whilst wearing gloves or other protective clothing. On aircraft lighting can be used where there are light limitations for critical tasks such as those of critical importance in achieving the necessary standards of maintenance performance to achieve these objectives.

It is particularly important that the design of a system or component does not infringe normal maintenance practices and the reasonable expectations of the maintainer based on training and experience. Maintainers might reasonably expect, for example, that, on a dial, values will increase clockwise.

Understand inspection; design solutions that consider the physical environment in which maintenance is conducted can reduce the potentially negative impact that it can have on maintenance performance.

26.6.3 Protect the Aircraft and Protect the Maintainer

Design solutions can actively influence both the impact that the maintainer has on the aircraft (*e.g.*, through maintenance error or routine violation of procedures) and the impact that the aircraft has on the maintainer (*e.g.*, through the health and safety effects of aircraft design).

Examples of design features that are tolerant to the consequences of maintenance error or resistant to the effects of maintenance activity and maintenance environment include:

- Designing out safety critical maintenance tasks;
- Items physically impossible to assemble or install incorrectly;
- Staggered part positions;
- Partial operation or redundancy;
- Shaped switch tops, display formats, direction of operation, *etc.*; and
- Warning lights and illuminated test points.

Examples of design considerations to protect maintenance personnel from risks, hazards, incidents, injuries or illnesses include:

- Electrical isolation and protection from high voltages;
- Adequate circuit breakers and fuses;
- Rounded corners and edges;
- Warning labels;
- Hot areas shielded and labelled; and
- Hazardous substances and radiation not emitted.

Protecting the maintainer is important not only from a health and safety perspective – demands placed on the maintainer that can be potentially injurious can also lead to the occurrence of maintenance error.

Design can place undue physical stresses on the maintainer. The maintainer may be required to wear cumbersome protective equipment to work in particular areas of the aircraft such as fuel tanks. The fatigue that can result could generate error. Other stressing design characteristics are those that, for example, involve inadequate lighting, vibration or noise, undue strength requirements for maintenance activities, unusual positions in which to carry out maintenance, or proximity of hot surfaces. A maintainer who must work close to heat generating components in a humid environment may rapidly lose body fluid, through perspiration as a result of increasing body temperature, which will seriously affect the ability to function correctly. If working close to a hot component, the maintainer must continuously avoid being burned whilst undertaking the maintenance task. The presence of such psychological and physical stressors can potentially lead to error.

Example 26.4: The Boeing 777 Refueling Panel (Sabbagh, 1996).

Boeing didn't think of the fact that existing fuel stands only reached a certain height to fuel under the wings of the airplane. The 747 was about as high as the

fuel stands could go to reach that fuelling panel, and the panel designed on the 777 was 31 inches higher than the 747.

Fuellers got very upset. “Have you ever fuelled an airplane in a high wind at O’Hare?” they said: “it’s really uncomfortable.”

To go any higher without additional stability would be a safety issue. Unless the operators hired personnel who are eight feet tall it wouldn’t work.

Boeing agreed to move the panel down the wing, closer to the fuselage, and, because the wing is slanted up, by moving it inboard it also came closer to the ground - within six inches of reaching the panel. Safety specialists allowed a stool to be put on the top of the fuelling platform to reach the panel.

26.6.4 Avoid Complexity of Maintenance Tasks

The design of a system or component will impact upon both the cognitive (thinking) and physical (doing) demands of the maintenance task. Complexity in design can generate complex maintenance tasks that are difficult to understand and difficult to do.

However, the avoidance of complexity in design need not compromise or constrain the technical design solution. The design principle is concerned with the effect that the design has on the maintenance task – an advanced design solution does not necessarily generate complexity in maintenance.

Example 26.5: Airbus A320 Flap Rotary Actuator (Airbus, 2005).

There are four rotary actuators on each wing of the A320. The function of these actuators is to translate the rotary motion of the flap drive shaft into movement of the flaps. Following flap lock events, it was reported in several cases that the flap rotary actuators had recently been removed for re-greasing.

Investigation revealed that, during accomplishment of removal or installation slight mis-rigging in the flap transmission had been induced. This was found to be a contributing factor in the reported flap locks. Existing flap rotary actuators filled with grease needed removal for re-greasing approximately every 5 years. A new type of actuator introduced is filled with semi fluid and is serviceable on the wing.

The design solution simplified the maintenance task by eliminating the need for removal/installation of the actuators, thereby removing the opportunity for mis-rigging.

26.6.5 Enable Adequate Maintenance Access

Accessibility means having adequate visual and physical access to perform maintenance safely and effectively. Adequate physical and visual access is needed not only for repair, replacement, servicing, and lubrication but also for troubleshooting, checking and inspection.

Examples of physical access considerations include:

- Adequate access to frequent maintenance areas;
- Openings of adequate size;

- Avoidance of the need to remove a large numbers of components, fittings, *etc.*, to reach a component;
- Replacement of components with the least amount of handling; and
- Workspace for manipulative tasks, body and tools positions and movements;

Examples of visual access considerations include:

- Avoidance of unnecessary obstructions to the maintainer's line of sight; and
- Lighting level and direction.

Some components by their function or requirements have to be located in poorly accessible areas – a design solution in such cases might be the use of integrated access platforms or other aids to access.

Example 26.6: B-1B Engine Visual Access (Worm, 1997).

Each engine on the B-1B bomber has an accessory drive gearbox (ADG). A hinged access door with four thumb latches is provided on each compartment panel for servicing. The access door permits checking of the ADG oil without having to remove the compartment panel. However, the oil level sight gauge requires line-of-sight reading. Because of the way it is installed, the gauge cannot be read through the access door, even with an inspection mirror. The entire compartment panel, secured with 63 fasteners, must be removed just to see if oil servicing is needed.

26.6.6 Positively Standardise and Positively Differentiate

Aircraft maintenance tasks are largely repetitive and standardised. Maintainers rely on pattern recognitions that are determined by their training and experience to identify system and component type properties and the form of the maintenance tasks that are required.

Commonality in design enables such pattern recognition and enhances maintenance performance. If, for example, a part has commonality in function and properties (and, of course, fully meets all requirements of the design specification) then it makes sense from the maintenance perspective to use common parts.

Similar systems or components with variations in configuration can reduce the effectiveness of maintenance and can be a cause of maintenance error. Reinforcement of pattern recognition can also be applied to commonality in maintenance activities.

If a part does not have commonality with the function and properties of other parts then it makes sense from the maintenance perspective to make the differences obvious. This will provide a clear and unambiguous signal to the maintainer that there are differences in maintenance actions.

Example 26.7: Boeing 777 Door Hinges (Sabbagh, 1996).

Early in the design process it was realized that there were three separate hinges that are complex parts. In addition, if the hinge came into the door at a different place on each door all the mating parts would be different. It was recognized early on that the key to making all the parts common was to make the hinge common, notwithstanding the fact that the shape of the body was different.

As a result, not only is the hinge common but so is the complete mechanism. Indeed, 98% of all the mechanism of the door is common.

26.6.7 Build Error Detection into the Maintenance Process

Design solutions can assist in the detection of maintenance error before aircraft dispatch. Design can determine how maintenance error is detected and by whom. Ideally, maintenance error should be detected before the aircraft is handed back to service after maintenance has been completed. In practice, however, the flight crew often detects error either before take off or, worse, in flight.

Mechanisms to detect error may include built-in tests, functional tests, illuminated test points, functionally grouped tests or warning lights, but equally they can be very simple, such as the use of physical indicators.

Ambiguous, difficult, complex or lengthy means to detect a maintenance error can affect the likelihood of detection being successful. Detection means should ensure that the maintenance error is evident under all maintenance conditions, easy and quick to detect, and detected before flight.

Example 26.8: JSF Landing Gear Sensors (DSI International, 2004).

The Joint Strike Fighter team has broken new ground by the use of landing gear sensors purely on the basis of improving maintenance performance.

Landing gear present many maintenance problems – one particular problem is measurement of the amount of hydraulic fluid by observation. This maintenance task has led to damaged landing gear due to overfilling.

The JSF programme, on the recommendation of its prognostics team, has agreed to embed sensors in the landing gear in order to report the exact level of hydraulic fluid, and in doing so has avoided maintenance error and saved cost.

26.7 Conclusion

There is a growing awareness of the vital role that design has to play in influencing maintenance performance and, more specifically, the avoidance or mitigation of maintenance error and its negative effects on safe and effective maintenance activity.

The maintainer interacts with aircraft systems and components through maintenance tasks that are generated by design characteristics. Design will determine the characteristics of the maintenance task and influence the possibility

of error occurring – it will also determine the possibility for error avoidance and tolerance. This chapter has described an analytical approach and general design principles that can be practically adopted and implemented to develop practicable solutions that address reasonably foreseeable maintenance errors.

The methodology and principles have been developed from extensive investigation of maintenance error, its causes and consequences specifically to enable the designer to consider the impact of physical design on the behavior of the maintainer.

The approach taken is deliberately not intended to prescribe design practice, to teach designers how to design, or to advocate further constraints to the design process but rather to add a vitally important dimension to existing knowledge and skills that will enhance maintenance performance and aviation safety.

References

- Airbus (2005) FAST Technical Magazine A320 Special, May.
- Boeing Commercial Airplane Group, (2003) Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations 1959–2002 Air Safety, May.
- CBC, (2005) CBC Canada News Updated 11 March.
- Courteney H (2001) Human Centred Design for Maintenance, UK Civil Aviation Authority, 15th HFIAM Symposium.
- DSI International (2004) A Prognosis Sensor Victory on the Joint Strike Fighter (JSF), November.
- F-16.net (2005) 23 August.
- Flight International (2003) Safety Review.
- Goglia J (2000) Unpublished statements at the 14th Human Factors in Aviation Maintenance Symposium, Vancouver, 2000 and Advances in Aviation Safety Conference, Daytona Beach, 2000.
- Graeber RC, and Marx DA (1993) Reducing Human Error in Aviation Maintenance Operations, Flight Safety Foundation 46th Annual International Air Safety Seminar, Kuala Lumpur, Malaysia.
- Hessburg J (2001) Air Carrier MRO Handbook, McGraw Hill.
- Marx D (1998) Learning from our Mistakes: A Review of Maintenance Error Investigation and Analysis Systems, FAA, January.
- MEMS-MEDA (2003) Seminar, May 2003.
- National Transportation Safety Board, United States of America, (2002) Aircraft Accident Report NTSB/AAR-03/02.
- Reason J (1997) Managing the Risks of Organisational Accidents, Ashgate.
- Royal Aeronautical Society, (2000) The Business Case for Human Factors Programmes in Aviation Maintenance, Royal Aeronautical Society Brainstorming Session, Gatwick, 8 December.
- Sabbagh K (1996) 21st Century Jet – The Making of the Boeing 777, Pan Books.
- Sears RL (1993) A New Look at Accident Contributions and the Implications of Operational Training Programmes (Unpublished Report) in Graeber RC and Marx D, Reducing Human Error in Aviation Maintenance Operations, Flight Safety Foundation Annual International Air Safety Seminar.
- Worm CM (1997) The Real World – A Maintainer's View, Proceedings IEEE Reliability and Maintainability Symposium, IEEE.

Index

- accident, 616
 - causation, 626
 - causes, 623
 - description, 624
 - examples, 622
 - industry examples, 624
 - prevention, 636
 - rate, 620
 - reasons, 621
 - severity, 620
- activity-based costing, 128
- Agency Theory, 474
- analysis of error, 719, 723, 726
 - worksheet, 719
- analysis of suspended data, 70
- ARMA, 171
- authority, 4, 10, 13, 14
- autonomous maintenance, 445
- autonomous work teams, 440
- autoregressive, 171
- availability, 192, 193, 196, 207, 304,
314, 315, 317, 321
- availability model, 138
- backlog, 108
- base warranty, 465
- Bayesian parameter estimation, 347
- budget
 - breakdown maintenance, 120
 - capital, 119
 - cash, 119
 - committee, 120
 - cycle, 120
 - definition, 116
 - performance, 117
 - planned maintenance, 120
 - template, 121
 - variances, 117
- budgetary process, 118
- budgeting, 107
- capacity planning, 6
 - definition, 175
 - formulation, 177
 - mathematical programming, 180
 - process, 175
 - queueing models, 183
 - simulation, 187
 - techniques, 176
- centralized maintenance, 7, 9
- centralized stock, 215
- change management, 20
- CMMS, 108, 422, 442
- co-generation system, 364, 387, 388, 392
- combined maintenance task, 579
- condition monitoring, 342, 500, 503,
506, 520, 524
- condition-based maintenance, 500
- control
 - actions, 348, 351
 - adaptive, 342, 348, 351, 356
 - algorithms, 350
 - cost, 109
 - parameters, 350
 - process, 96, 339, 350, 356
 - signals, 351
 - statistical, 350
 - steps, 98
 - strategy, 350, 351
 - structure, 97

- tools, 107
- control barriers, 84
- corrective deferred major, 174
- corrective deferred minor, 173
- corrective emergency, 173
- corrective maintenance, 363
- cost variance, 129
- CPM, 253, 256
- Croston method, 199, 200
- cumulative damage, 364, 388, 389, 392
- cycle average method, 169
- data
 - censored, 63
 - complete, 63
- decentralized maintenance, 7, 9
- decentralized system, 214
- decision support system, 506
- Decision Support System, 518
- degradation process, 312, 321, 324, 326, 328
- Delphi method, 160
- Deming's 14 points, 650, 651, 652
- detection, 77
- detection evaluation criteria, 81
- detection signal, 347
- diagnosis and prognosis, 512
- discounted Markov decision problem, 270
- discounted Markov decision process, 270, 273
- distribution
 - binomial, 54
 - definition, 50
 - exponential, 60
 - lognormal, 58
 - normal, 56
 - Poisson, 55
 - Weibull, 61
- duration of maintenance task, 566
- dynamic programming principle, 276
- e-maintenance application, 524, 536
- e-maintenance solution, 503, 507, 508, 510, 528, 529, 530, 536, 537
- engineering controls, 639
- environmental performance, 656
- equipment history, 409
- ERP, 260
- estimation method
 - least squares, 67
 - maximum likelihood, 69
 - probability plotting, 64
- European Agency for Safety and Health at Work, 643
- exponential smoothing
 - double, 166
 - simple, 165
- extended warranty, 467
- factor charting, 84
- failure detection, 347
- failure effect, 77
- failure mode, 77
- failure modes, 398, 400, 405
- failure rate, 52
- fault diagnosis, 356
 - techniques, 350
- fault diagnosis and detection, 349
- Fault Tree Analysis, 638
- feedback, 99
- FEMA, 410
- FIMS, 107
- FMEA, 638
- forecasting model
 - steps, 158
- forecasting techniques, 158
- free replacement warranty, 466
- full authority digital electronic control, 363
- full authority digital engine controller, 380, 381, 385
- functional block diagram, 407
- functional failure, 398, 405, 410
- functionality, 548
- functions, 6, 9, 15
- Gantt chart, 250
- hazard, 616
 - classes, 619
- hazard prevention, 638
- health monitoring, 537
- history file, 103
- Holt's method, 166
- human error, 712, 713, 718, 724, 726
 - consequence, 700
 - in maintenance, 625
 - influencing factors, 626
 - reasons, 701
 - reduction, 707
- human errors
 - categories, 625
 - classification, 700

- human performance reliability, 699
- imperfect repair, 468
- incentives, 4, 13
- indices
 - maintenance, 96
 - production, 96
- industrial hazard avoidance, 637
- in-house maintenance, 4, 6, 7
- inspection, 365, 366, 367, 370, 372, 482
 - alarm threshold, 323
 - basic model, 305
 - conditional, 322
 - degradation process, 324
 - example, 493
 - imperfect, 485
 - modeling, 483
 - multi-component system, 318
 - policies, 311
 - perfect, 484
- intelligent maintenance system, 500, 506, 510, 511, 537
- International Labour Organization, 643
- inventory
 - control, 199
 - control policies, 201
 - models, 202
 - parameters, 201
 - policies, 205
 - pooling, 214, 215
 - spare parts, 213
- job card, 102
- job cost sheet, 125
- joint optimization of maintenance and inventory, 205
- joint replenishment, 214
 - spare parts, 216
- labor relations, 5, 14
- LCC, 108
- leadership, 4, 13
- life cycle costing, 634
- likelihood function, 487, 489, 491, 492, 494
- logic tree analysis, 412
- maintainability
 - analysis, 561
 - concept, 549, 550
 - definition, 551
 - design issues, 596
 - design review, 587
 - desirable practices, 559
 - economic impact, 554
 - engineering, 572
 - engineering management function, 592, 593
 - engineering predictions, 577
 - function, 565, 567, 574, 582, 586
 - function derivation, 580
 - impact of system effectiveness, 552
 - impact on safety, 555
 - lessons learned, 603
 - measures, 568, 571, 581, 583
 - parametric approach, 572
 - statistics, 561, 572
 - undesirable practices, 558
- maintainability block diagram, 577
- maintainability data
 - distribution approach, 572
- maintenance error
 - causes, 724
 - definition, 712
 - design, 717
 - detection, 719
 - examples, 713, 714, 715, 722, 734
 - mitigation, 727, 731
 - reasons, 712
 - safety impact, 713
- maintenance error analysis, 702
- maintenance models, 363
- maintenance performance, 138, 146
- maintenance philosophies, 340
- maintenance prevention, 421, 449
- maintenance strategy, 138, 142, 152, 243
- maintenance system audit, 655
- Markov decision problem, 279
- Markovian jumps, 275
- maximum likelihood method, 135
- mean absolute deviation, 172
- mean absolute percent error, 172
- mean duration of maintenance task, 566, 575
- mean squared error, 172
- mean time between failure, 53
- Mean Time Between Maintenance, 569
- Mean Time Between Replacements, 569
- median rank, 65
- missile, 364, 365, 373
- model based approaches, 344
- moral hazard, 474
- moving average, 162
 - model, 171

- simple, 162
 - weighted, 162
- MPIs
 - auto industry, 38, 39
 - categories, 29
 - definition, 24
 - development, 25
 - oil industry, 35
 - process industry, 37
 - selection, 26
 - specifying, 26
 - testing, 29
 - utility industry, 37
- MPM
 - implementation, 29, 31
 - issues, 25
 - multi-criteria, 31
 - oil industry, 35
 - process industry, 37
 - system, 27
- MTBM. See Mean Time Between Maintenance
- MTBR. See Mean Time Between Replacements
- neural networks, 344, 347, 350, 353
- occupational safety, 613, 620, 632, 642, 643
- occupational stressors, 697
- occurrence, 77
- occurrence evaluation criteria, 79
- OEE, 110, 426
- OPTIMAIN, 134, 141, 150, 152
- optimal control problem, 268
- organization structure, 3, 6, 15
 - design, 4, 8, 15
 - hybrid, 7, 10
 - matrix, 10, 12
- organizational change, 8
- OSHA, 642
- outsourcing, 4, 6, 7, 15, 242
- parameter estimation, 64, 345
- parity space relations, 346
- patterns, 159
- percentage restoration time, 574
- percentual duration of maintenance task, 566
- performance indicators, 22, 33, 34, 36, 661
- personal protective clothing and/or equipment, 640
- PERT, 253, 258
- phased array radar, 364, 373, 374, 377
- planner qualifications, 247
- planning and forecasting, 98
- planning horizon, 237
- planning procedure, 246
- planning sheet, 248
- PM task selection, 411
- PPE. See personal protective clothing and/or equipment
- predictive maintenance, 400, 420
- preventive maintenance, 363, 420
- productive maintenance, 421
- productivity
 - measurement, 21, 22
 - measurement factors, 23, 24
- prognostic approach, 502
- prognostic systems, 348, 356
- proportional hazards modeling, 347
- quality policy, 231
- RCM, 342, 421, 635
- RCM process, 398, 400, 411
- regression method, 134, 135
- regression models, 163
- reliability estimation, 134, 152
- reliability function, 52
- reliability model estimation, 138
- repeat jobs, 109
- replacement at *pro-rata* cost, 466
- residual generation, 345
- responsibility, 5, 6, 13
- risk pooling, 215
- root cause analysis, 638
- run to failure, 400
- safe maintenance practices, 639
- safety culture, 641
- safety legislation, 642
- safety monitoring, 232
- safety performance indicators, 645
- safety performance measurement, 644
- safety policy, 231
- sales force composite, 160
- scheduling
 - definition, 237, 247
 - elements, 249
 - objectives, 239
 - procedure, 249

- seasonal factors, 169
- sequential maintenance task, 579
- service delivery, 471
- service level, 198, 200, 215
- severity, 77
- severity evaluation criteria, 80
- signal-based FDI, 346
- simulation model, 134, 143, 144, 150, 151
- spare part inventories, 473
- spare parts
 - centralization, 215
 - characteristics, 191
 - classification, 193
 - forecasting, 198
 - identification, 192
 - joint ordering, 216
 - kit, 192
 - maintenance, 212
 - optimal quantity, 195, 202
 - quantity, 193
 - reconditioned, 210
 - replacement, 207
 - virtually centralized, 214
- spare parts management, 4, 10
- standby systems, 319
- state trajectories, 276
- stationarity, 159
- stimuli, 616
- strategic planning process, 243
- supervision, 4, 5, 7, 13
- sustainability
 - business, 672
 - compliance, 671, 676, 687
 - drivers, 674
 - issues, 673
 - maintenance impact, 679
 - performance, 678
 - risk, 687
- sustainability performance measurement, 670
- sustainable asset maintenance, 667
- system boundary, 407
- system function, 398, 400, 401, 405, 407, 409, 410
- system interfaces, 408
- TAM phases, 225
- time-frequency analysis, 347
- TPM, 635, 650
 - benchmark indices, 451
 - definition, 418
 - implementation, 429
 - KPI, 451
 - lean practices, 452
 - methodology, 435
 - need, 423
 - obstacles, 453
 - safety, 449
 - success factors, 456
 - tools, 425
- training, 4, 7, 10, 14
- types of censored data, 64
- types of FMEA, 79
- vibration monitoring, 350
- visual workplace, 442
- WAN, 507, 510
- warranty and maintenance, 467
- warranty logistics, 472
- warranty servicing model, 468
- watchdog agent, 510, 512, 514, 518, 525, 532, 536
- wear, 323
- web-based solutions, 505
- wide-area networks, 505
- work measurement, 107
- work order
 - flow, 105
 - form, 100
 - information, 101
 - planning, 98
 - system, 99
- work package, 227
- working relations
 - chain of command, 4
 - delegation of authority, 4