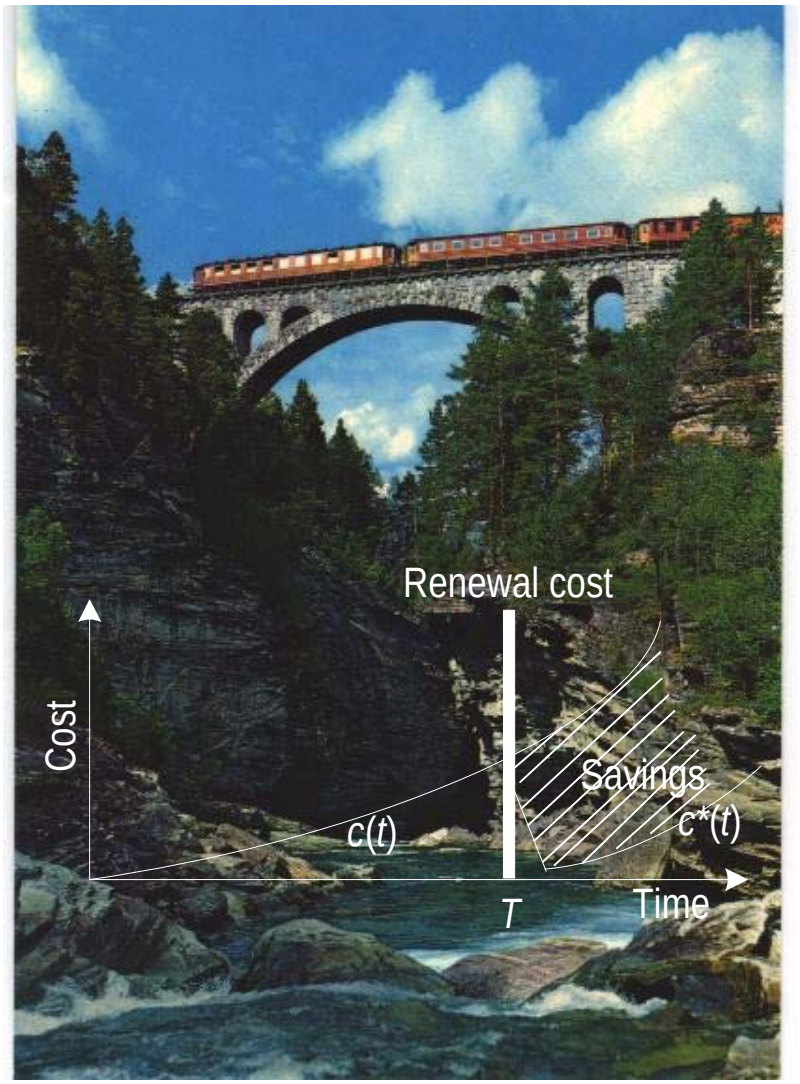
Jørn Vatn

NTNU

Veien frem til “World Class Maintenance”:

Maintenance Optimisation



PREFACE

This course material has been developed for a course in railway maintenance optimisation arranged by the Norwegian University of Science and Technology (NTNU). A future plan is to develop this material into a textbook on the topic.

At the time being, most examples in this report are taken from railway applications, and special acknowledge is made to:

- The Norwegian National Railway Administration (JBV) for valuable input I have got during my work at the project “Vedlikehold av jernbanenettet”.
- The European Union for economical support during the ProM@in project.

Even if most examples relates to railway applications, the presentation is rather general, and the methods and models could also be used in other industries.

Jørn Vatn

Trondheim, November 2007

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

This course deals with maintenance optimisation within railway application. With maintenance we understand "the combination of all technical and administrative actions, including supervision actions, intended to retain an item in, or restore to, a state in which it can perform a required function". With maintenance optimisation we understand "balancing the cost and benefit of maintenance". There are many aspects of maintenance optimisation, and some of these are:

- Deciding the amount of preventive maintenance (i.e. choosing maintenance intervals).
- Deciding whether to do first line maintenance (on the cite), or depot maintenance.
- Choosing the right number of spare parts in stock.
- Preparedness with respect to corrective maintenance.
- Time of renewal.
- Grouping of maintenance activities.

The main focus in this course will be on optimising preventive maintenance intervals and time of renewal. Other aspects will however also be treated to some extent.

Exercise 1

Identify areas within your organisation where maintenance optimisation is of interest. □

1.1 The bath tub curve and the failure/hazard rate

Most methods and approaches to maintenance analysis involve the concept of *hazard rate*. Very often the hazard rate shows a bath tub like behaviour as illustrated in Figure 1. The hazard rate defines the probability that an item will fail in a small time interval from time t to $t + \Delta t$ given that the item has survived up to time t.

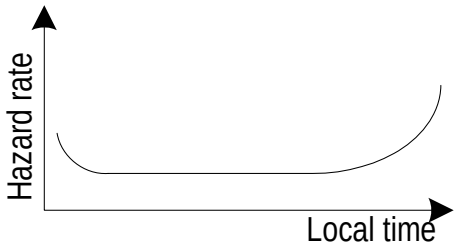


Figure 1 Bath tub or hazard rate function

In Figure 1 we have used the word "local time" to emphasise the fact that time is relative to the last failure (or maintenance point), rather than to the global system time. The bath tub curve indicates that the number of failures will be reduced if the component is replaced or maintained before we run into the right part of the curve. There exists also another bath tub curve related to the *global* system time as shown in Figure 2 where we also have illustrated the local bath tub curves.

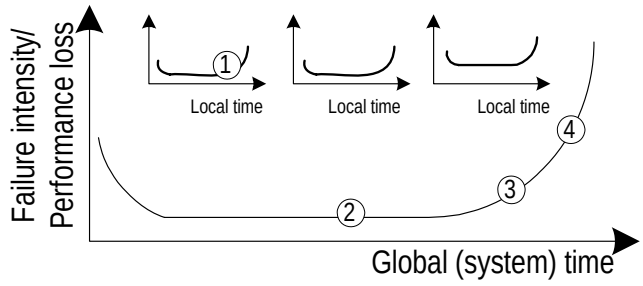


Figure 2 Global system time

As an example, consider a signalling system with lights, logic's, relays etc. The local time (time horizon 1 to 5 years) applies to the light bulbs, the relays etc, whereas the global time (time horizon 30-60 years) applies when the entire signalling system is considered. Note further that on the *y*-axis the dimension is *failure intensity*, or performance loss. This reflects that the important issue now is the number of failures per unit time, or generally loss of performance, independent of what has happened up to time *t*.

In Figure 2 we have also identified the numbers ①, ②, ③, and ④, where the following maintenance situations apply:

① Component maintenance, related to the explicit failure modes of a component. FMEA[1] and RCM[2] analysis is relevant. A typical question is "*when should we on a preventive basis replace light bulbs in the signalling system?*"

② Life extension maintenance. The idea here is to carry out maintenance that prolongs the life length of the system. A typical example is "*rail grinding to extend the life length of rails*".

③ Maintenance carried out in order to improve performance, but not renewal. A typical example is "*adding ballast to pumping sections to improve track quality and reduce the need for track adjustment*".

④ Complete renewal of major railway components or systems.

1.2 Preventive maintenance and RCM

With preventive maintenance (PM) we understand "the maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of an item" (EN 13306). There exist several approaches to determine a preventive maintenance program. A concept that is becoming more and more popular is the concept of Reliability Centred Maintenance (RCM). RCM is "a systematic consideration of system functions, the way functions can fail, and a priority–based consideration of safety and economics that identifies applicable and effective PM tasks".

An RCM analysis is usually conducted as a pure qualitative analysis with focus on identifying appropriate maintenance tasks. However, the RCM methodology does not give support for quantitative assessment in terms of e.g. interval optimisation. In this course we will present the framework for optimising maintenance interval as well.

[1] Failure Mode and Effect Analysis
[2] Reliability Centred Maintenance

The strength of RCM is its systematic approach to consider all system functions, and set up appropriate maintenance task for these functions. On the other hand, RCM is not a methodology that could be used to define a renewal strategy (see ④ in Figure 2). To determine optimal renewal strategies we will in this course work with Life Cycle Cost modelling (LCC).

1.3 Renewal and Life Cycle Cost

When the system deteriorates to a certain level, traditional preventive maintenance activities could not bring the system to a satisfactory state, and renewal of the entire system, or part of the system is required. However, the cost of renewal is often very large, and we need formalised methods to determine when to perform renewal. In this course we will present methods for optimum renewal strategies based on LCC modelling. The following dimensions are included in the LCC model: *i*) safety costs, *ii*) punctuality costs, *iii*) maintenance & operational costs, *iv*) cost due to increased residual life length, and *v*) project costs. The LCC models apply to ②, ③, and ④ in Figure 2.

1.4 Reliability modelling

Formalised maintenance optimisation models rely on system reliability models. These are models that express the system (reliability) performance as a function of component performance. Further the component performance is expressed in terms of component reliability models. Therefore we will in this course also present a toolkit of standard reliability models. These models are:

- Reliability block diagram (RBD) and structure functions.
- Fault tree analysis (FTA).
- Event tree analysis (ETA).
- Markov analysis.
- Failure Mode and Effect Analysis (FMEA/FMECA).

We will also present an introduction to probability theory and common probability distributions.

1.5 Basic maintenance models

Within maintenance optimisation literature it is common to present some basic models such as the Age Replacement Policy (ARP) model, the Block Replacement Model (BRP) and the Minimal Repair Policy (MRP). Such models were introduced by Barlow and Hunter (1960) and have later been generalised in several ways, see e.g. Block *et. al.* 1988, Aven and Bergman (1986), and Dekker (1992). There exists also several major (review) articles in this area, e.g. Pierskalla and Voelker (1979), Valdez Flores and Feldman (1989), Cho and Parlar (1991) and Wang (2002).

In this presentation we will not discuss these standard models in detail. Our approach aims at establishing what we denote the “effective failure rate”. This effective failure rate is the failure rate we would experience if we (preventive) maintain a component at a given level, and mathematically we let $\lambda = \lambda(\tau)$, where λ is the effective failure rate, and τ is the maintenance interval. Now there is two challenges, first we want to establish the relation $\lambda =$

$\lambda(\tau)$ depending on the (component) failure model we are working with, then next, we need to specify a cost model to optimise. The cost model will generally involve system models as fault tree analysis, Markov analysis etc. This enables us to find the optimum maintenance intervals in a two step procedure. Note also that when we use $\lambda = \lambda(\tau)$ in the system models we then assume a "constant failure rate" which of course is an approximation for ageing components. However, if the component is maintained, such an approximation could be reasonable.

1.6 Introductory example

Consider a component for which the effective failure rate is given by $\lambda = \lambda(\tau) = \tau/100$, where τ is the maintenance interval. Assume that the cost of a component failure is $CM_{Cost} = 10$ (corrective maintenance cost including loss of production during the repair period). Further let $PM_{Cost} = 1$ is the cost per preventive maintenance action carried out at intervals of length τ. The total cost per unit time is then given by:

$$C(\tau) = PM_{Cost} / \tau + CM_{Cost} \times \lambda(\tau) = 1 / \tau + \tau/10 \qquad (1)$$

The interval that mimeses the cost could easily be found by differentiation, but we could also graphically plot the cost as a function of the maintenance interval (τ). The result is shown in Figure 3, and we see that the optimum maintenance interval is $\tau = 3$. Very often such a graphical method is sufficient.

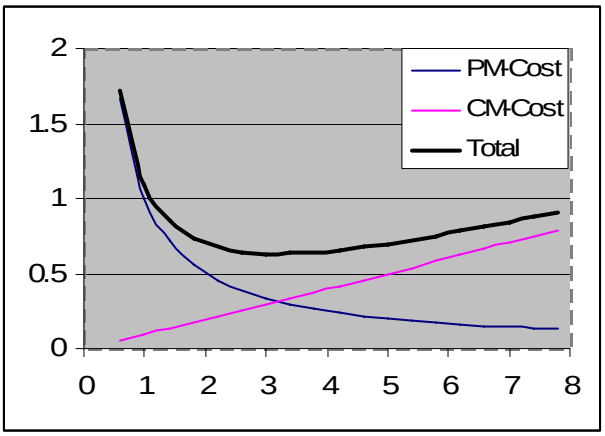


Figure 3 Optimising maintenance interval

1.7 Utility programs

In order to perform the calculations in real situations it is necessary to have access to computerised tools. Thorough this report we have made use of simple MS Excel utility programs. These programs could be downloaded from www.ntnu.no/ross. Currently the following programs are available:

RDBUtil.xls: Utility program for reliability block diagram.
WeibullRenewal.xls: Program for calculation of renewal function in the Weibull distribution.
MAKROV.xls: Utility program for Markov analysis.
MaintOpUtil.xls: Simple program for optimisation of maintenance intervals.
PFCalc.xls: Program for calculation of "effective failure rate" in the "PF-situation".

1.8 Notation and definitions

α	Ageing parameter in the situation of increasing hazard rate, $z(t)$
λ	Failure rate in the situation of constant hazard rate, $z(t)$
$\lambda_A(\tau)$	Effective failure rate for an aging component that is replaced after failure, and preventive renewed or replaced at maintenance interval τ
$\lambda_E(\tau)$	Effective failure in the general situation, i.e. when the component is maintained at intervals of length τ.
λ_N	The naked failure rate, i.e. the expected number of failures per unit time when the component is not maintained.
λ_i	Failure rates in a Markov model
$\phi(\boldsymbol{x},t)$	Structure function of a system, 1 if the system is functioning at time t, 0 otherwise
τ	Maintenance interval in the general situation
τ_A	Interval for preventive replacement/overhaul for ageing components
τ_{HF}	Interval for functional test (hidden function)
τ_{PF}	Interval for condition monitoring (PF situation, Failure progression)
μ_i	Repair rates in a Markov model
ν_j	Visiting frequency, i.e. number of visits to state j per unit time
$\{RC(t)\}$	Portfolio cost of renewals without a maintenance project
$\{RC^*(t)\}$	Portfolio cost of renewals with a maintenance project
A	Transition matrix in a Markov mode
$C(\tau)$	Total cost per unit time when the component is maintained at intervals of length τ
$c(t)$	Variable cost in Renewal optimisation
$c^*(t)$	Reduced variable cost if renewal or maintenance project is executed
$C_{CM}(\tau)$	Corrective maintenance cost per unit time when the component is maintained at intervals of length τ
CM	Corrective maintenance, i.e. maintenance carried out after a failure to restore the function of an item.
CMMS	computerized maintenance management system
$C_{PM}(\tau)$	Preventive maintenance cost per unit time when the component is maintained at intervals of length τ
$C_S(\tau)$	Safety cost per unit time when the component is maintained at intervals of length τ
E_{PF}	Expected, or mean value of the PF interval
ETA	Event Tree Analysis
F_0	Frequency of the TOP event in a fault tree
Failure	Termination of ability to perform the required function
Fault	The state that the required function could not be performed
f_D	Demand rate for which the hidden function is demanded
FOM	Force of mortality, $h_X(x) = \frac{f_X(x)}{1 - F_X(x)}$. FOM is identical to hazard rate
f_P	Frequency of "potential failures", i.e. the number of "P"s in the "PF interval"
FTA	Fault Tree Analysis
$F_X(x)$	Distribution function, or life time distribution , $F_X(x) = Pr(X \leq x)$

$f_X(x)$	Probability density function, $f_X(x) = \frac{dF_X(x)}{dx}$
$h(\mathbf{p},t)$	System reliability, the probability that the system is functioning at time t, as a function of the component reliabilities p = [p_1, p_2,…]
HPP	The Homogeneous Poisson Process
$I^B(i)$	Birnbaums measure of reliability importance, $I^B(i) = \partial\, h(\mathbf{p})/\, \partial p_i$
$L(\boldsymbol{\theta},\mathbf{t})$	Likelihood function, used to estimate life time parameters
LCC	Life Cycle Cost
MDT	Mean Down Time
MTBF	Mean Time Between Failures, MTBF = MTTF + MDT
MTTF	Mean time to failure. We use the index N to indicate the "naked" MTTF if no maintenance is carried out ($MTTF_N$), and the index E to indicate the effective failure rate if maintenance is carried out. $MTTF_E$ will then be a function of the maintenance interval.
MTTR	Mean Time To Repair
$N(t)$	Cumulative number of failures from 0 to t
NHPP	The Non Homogeneous Poisson Process
P	Steady state probabilities in a Markov model, **P** = [P_0,P_1,…]
PF-interval	Time from a potential failure (P) is detected until a failure (F) occurs
PM	Preventive maintenance, i.e. the maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of an item
PFD	Probability of failure on demand
$Q_0(t)$	Probability that the TOP event in a fault tree occur at time t, or system failure probability
$Q_{FP}(\tau)$	Probability of not detecting a "potential" failure in due time in the situation of observable failure progression
$q_i(t)$	Probability that component i does not function at time t, or probability that a basic event has occurred at time t in a fault tree
$Q_M(\tau)$	Probability that the maintained barrier does not function as intended when maintained at intervals of length τ
R	Interest rent
$R(x)$	Survival probability, $R(x) = Pr(X>x) = 1\text{-}F_X(x)$
RAMS	Reliability, Availability, Maintainability and Safety
RBD	Reliability Block Diagram
RBI	Risk based inspection
RCM	Reliability Centred Maintenance
Renewal	Renewing of a system when preventive and corrective maintenance is not sufficient, or cost effective to ensure sufficient performance of a system
RLL	Residual Life Length, i.e. time until the system could not be operated any more (if noting is done)
RLL*	Residual Life Length if a maintenance project or renewal is conducted
ROCOF	Rate of OCcurrence Of Failures, ROCOF = $w(t) = \frac{dW(t)}{dt}$, $w(t)\Delta t \approx Pr$(Failure in (t, $t+\Delta t$))
RP	The Renewal Process
SD_{PF}	Standard deviation of the PF interval
T	Life time of a component, when life times are treated separately

TiS	Time in Service (total time the unit has been in service)
TTT	Total Time on Test
U	Unavailability
$W(t)$	Expected cumulative number of failures in 0 to t , $W(t) = \mathrm{E}[N(t)]$
$W(t)$	Renewal function, i.e. number of renewals in 0 to t if the unit is renewed after a failure
X	Life time of a component, when each component has several "life times", i.e. the component is replaced or renewed after a failure
$x(t)$	State variable of components, 1 if the components is functioning at time t, 0 otherwise
$z(t)$	Hazard rate, $z(t) = \frac{f_T(t)}{1 - F_T(t)}$. Same as Force Of Mortality, FOM

2. MAINTENANCE MANAGEMENT

In this chapter we will highlight important elements of maintenance management. The discussion take the maintenance management loop in Figure 4 as a starting point, and each box is discussed.

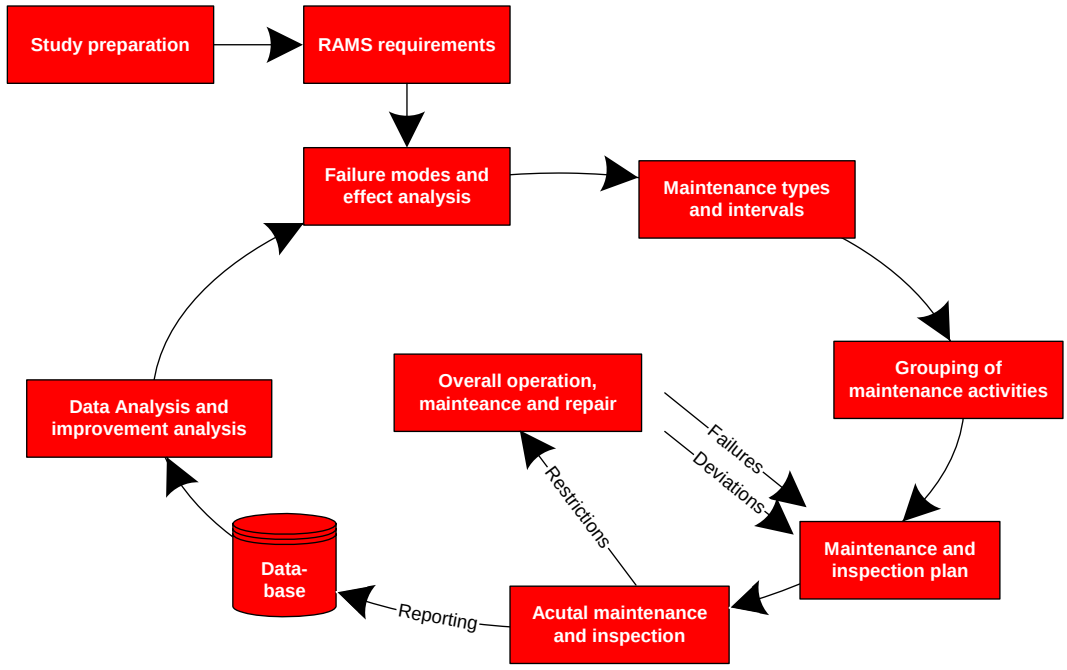


Figure 4 Maintenance Management Loop

2.1 Study preparation

It is important to define and clarify the objectives and the scope of the analysis. Requirements, policies, and acceptance criteria with respect to safety and environmental protection should be made visible as boundary conditions for the analysis. Further key persons should be identified, and a map of the maintenance organisation should be set up.

2.2 RAMS requirements

In order to set up an optimal maintenance and inspection plan the RAMS (Reliability, Availability, Maintainability and Safety) requirements has to be determined. The CENELEC standards EN 50126, EN 50 128 and ENV 50 129 are inputs to the RAMS requirements. Other inputs may be the “single fault principle” and control with safety critical functions.

2.3 Failure mode and effect analysis

Failure Mode, (Criticality) and Effects Analysis (FMCEA) was one of the first systematic techniques for failure analysis. It was developed by reliability engineers in the late 1950’s to determine problems that could arise from malfunctions of military systems.

A Failure Mode and Effects Analysis is often the first step in a systems reliability study. It involves reviewing as many components, assemblies and subsystems as possible to identify possible failure modes and the causes and effects of such failures. For each component, the failure modes and their resulting effects on the rest of the system are written onto a specific FMCEA form.

2.4 Maintenance and inspection types and intervals

The main objective of this step is to determine the type and frequencies of maintenance and inspection tasks. In principle each failure mode/failure cause in the FMEA should be combated by a maintenance task. The RCM logic of an RCM analysis will be a starting point for identifying relevant maintenance tasks. See Chapter 9 for an introduction to RCM. To determine optimal frequencies of maintenance tasks it is usually required to establish a cost model to optimize. Life cycle costing (LCC) will be a central part of such model. The use of so-called influence diagrams will very often help the communication between the analyst and maintenance engineers, economists etc.

2.5 Grouping of maintenance and inspection work

When the maintenance tasks are identified, and frequencies set it will usually be natural to group these activities into maintenance packages, each package describing what to do, and when to do it. It is a challenge to establish such an optimal grouping strategy.

2.6 Maintenance and inspection plan

A maintenance program shall be established, which includes written procedures for maintaining, testing, and repairing the various components within the railway system. Such a program is often implemented by a computerised maintenance management system (CMMS). A main task of the CMMS is to manage all work orders for preventive maintenance.

2.7 Failures need corrective maintenance

Failures represent technical component failures (e.g. rail breakage, defect breaks etc), and deviations (e.g. geometrical deviations of the track). Failures and deviations require repair, overhaul etc. Typically a work order for corrective maintenance (CM) is issued. The CMMS will also manage these work orders.

2.8 Reporting of result from maintenance and inspection

All maintenance work (functional testing, preventive maintenance, and corrective maintenance) shall be reported into an electronic maintenance database. The information to report depends on the type of maintenance work.

2.9 Deviations

The integrity of the track is to some extend ensured fulfilling some technical requirements related to e.g. geometry, rail profile, turnout distances etc. When some of these requirements are not fulfilled, it is necessary to issue corrective maintenance work.

2.10 Database

The database used in the maintenance management loop is a conceptual term. A RAMS database may be realised as a part of the CMMS. It is essential that the database system allows for storing the information necessary for a proper data analysis.

2.11 Data analysis and improvement analysis

It is essential that the scope of the data analysis is agreed upon. As a minimum the analysis should include:

- A proper failure cause analysis (FCA), or root cause analysis (RCA).
- Investigation into the failure reports to identify common cause problems (CCF).
- Updated reliability data that was used in quantitative risk analyses.
- Verification of assumption related to safety critical functions (SCF). For example there might be assumption about crack speed propagation in a rail within the inspection program. If this assumption does not hold, the inspection program should be changed accordingly. Key questions are *i*) Is the “failure rate” as expected? *ii*) is there a negative *trend* in the failure rate? *iii*) Is it possible to evaluate the failure propagation speed (P-F intervals)? *iv*) Is it experienced *new* failure modes that was not considered in the maintenance plan? *v*) Is it conditions related to the SCF that indicate ”wrong use”?, and finally, *vi*) Is it conditions that indicate that there are safety critical functions that we did not identify in the initial analysis.

The analysis group should also identify the need and relevance of:

- Reporting to the regulator;
- Feedback to the manufactures and vendors.

The results from the analysis are used to suggest improvement measures. The results could also be feed back into the risk analysis, e.g. did we experience higher failure rate than expected, and hence have to reconsider the situation. This may then results in changing the maintenance intervals.

2.12 Restrictions

When maintenance comes out of control (large backlog) it is important to initiate operational restrictions (e.g. closing the line, reducing speed etc). Restrictions will also be necessary when the track integrity is threatened by weather conditions such as rain, frost, snow, high temperature etc.

2.13 Overall operation, maintenance and repair

This “box” represent the physical or real activity required by a railway company “out there”. The results of this is obviously to fulfil the main objectives which is to run the trains, but in addition to this there will be failures, deviations, incident, accidents etc.

3. PROBABILITY THEORY

3.1 Basic probability notation

In this section basic elements of probability theory are reviewed. Readers familiar with probability theory can skip this section. Readers which are very unfamiliar with this topic are advised to read an introductionary textbook in probability theory.

Event

In order to define probability, we need to work with events. Let as an example A be the event that there is an operator error in a control room. This is written:

A = {operator error}

An event may occur, or not. We do not know that in advance prior to the experiment or a situation in the "real life".

Probability

When events are defined, the probability that the event occurs is of interest. Probability is denoted by Pr(·), i.e.

Pr(A) = Probability that A occur

The value of Pr(A) may be found by:

- Studying the *sample space*
- Analysing collected data
- Look up the value in data hand books
- "Expert judgement"

The *sample space* defines all possible events. As an example let A = {It is Sunday}, B = {It is Monday}, .. , G = {It is Saturday}. The sample space is then given by

S = {A,B,C,D,E,F,G}

So-called Venn diagrams are useful when we want to analyse subset of the sample space *S*. A rectangle represents the sample space, and closed curves such as a circle are used to represent subsets of the sample space as illustrated in Figure 5.

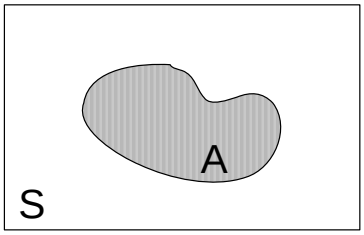
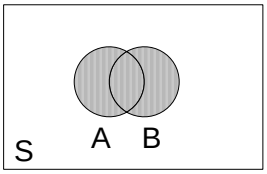


Figure 5 Venn diagram

In the following we will describe frequently used combinations of events:

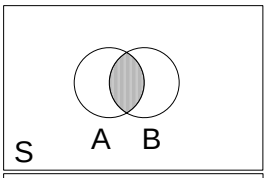
The *union* of two events A and B:

$A \cup B$ denotes the occurrence of A or B or (A and B).



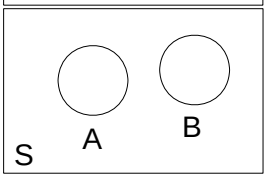
The *intersection* of two events A and B:

$A \cap B$ denotes the occurrence of both A and B.



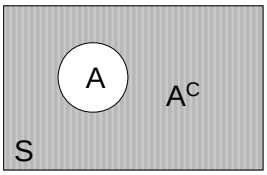
Disjoint events:

A and B are said to be *disjoint* if they can *not* occur simultaneously, i.e. $A \cap B = \emptyset$ = the empty set.



Complementary event:

The *complement* of an event A is all events in the sample space S except for A. The complement of an event is denoted by A^C.



Probability is a set function Pr() which maps events A_1, A_2,... in the sample space S to real numbers. The function Pr(·) can only take values in the interval from 0 to1, i.e. probabilities are greater or equal than 0, and less or equal 1.

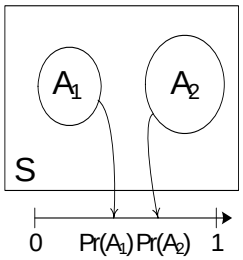


Figure 6 Mapping of events on the interval [0, 1]

Kolmogorov established the following axioms which all probability rules could be derived from:

1. $0 \leq Pr(A)$
2. $Pr(S) = 1$
3. If $A_1, A_2,...$ is a sequence of disjoint events we shall then have:
   $Pr(A_1 \cup A_2 \cup ...) = Pr(A_1) + Pr(A_2) + ...$

The axioms does not help us in establishing numerical values for $Pr(A_1)$, $Pr(A_2)$, etc. Historically two lines of thoughts have been established, the classical (frequentiest) and the Bayesian approach. In the classical thinking we introduce the concept of a random experiment, where $Pr(A_i)$ is the relative frequency with which A_i occurs. The probability could then interpreted as a property of the experiment, or a property of the world. By letting nature reveal itself by doing experiments, we could in principle establish all probabilities that are of interest. Within the Bayesian framework probabilities are interpreted as subjective

believe about whether A_i will occur or not. Probabilities is then not a property of the world, but rather a measure of the knowledge and understanding we have about a phenomenon.

Before we set up the basic rules for probability theory that we will need, we introduce the concepts of conditional probability and independent events.

Conditional probability
Pr(A|B) denotes the conditional probability that A will occur given that B has occurred.

Independent events
A and B are said to be *independent* if information about whether B has occurred does *not* influence the probability that A will occur, i.e. Pr(A|B) = Pr(A).

Rules for probability
The following calculation rules for probability can be used:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \quad (2)$$

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) \text{ (if A and B are independent)} \quad (3)$$

$$\Pr(A^C) = \Pr(\text{A does } \textit{not} \text{ occur}) = 1 - \Pr(A) \quad (4)$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad (5)$$

Example 3.1

Let A = {It is Sunday}
B = {It is between 6 and 8 pm)

A and B are independent but not disjoint.

We will find Pr(A ∩ B) and Pr(A ∪ B)

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) = \frac{1}{7} \times \frac{2}{24} = \frac{1}{84}$$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = \frac{1}{7} + \frac{2}{24} - \frac{1}{84} = \frac{9}{42}$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/84}{2/24} = \frac{1}{7}$$

□

3.2 The law of total probability

$A_1, A_2, \ldots, A_r$ is said to be a division of the sample space if the union of all A_i's covers the entire sample space, i.e. $A_1 \cup A_2 \cup \ldots \cup A_r = S$ and the A_i's are pair wise disjoint, i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$. An example is shown in Figure 7.

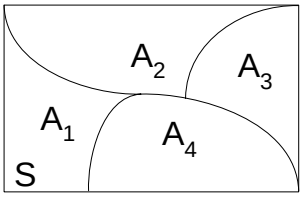


Figure 7 Division of the sample space

Let $A_1,A_2,\ldots,A_r$ represent a division of the sample space S, and let B be an arbitrary event in S. The law of total probability now states:

$$\Pr(B) = \sum_{i=1}^{r} \Pr(A_i) \times \Pr(B \mid A_i) \tag{6}$$

Example 3.2

A special component type is ordered from two suppliers A_1 and A_2. Experience has shown that components from supplier A_1 has a defect probability of 1%, whereas components from supplier A_2 has a defect probability of 2%. In average 70% of the components are provided by supplier A_1. Assume that all components are put on a common stock, and we are not able to trace the supplier for a component in the stock. A component is now fetched from the stock, and we will calculate the defect probability, Pr(B):

$$\begin{aligned}\Pr(B) = \sum_{i=1}^{r} \Pr(A_i) \cdot \Pr(B \mid A_i) &= \Pr(A_1) \cdot \Pr(B|A_1) + \Pr(A_2) \cdot \Pr(B|A_2) \\ &= 0.7 \cdot 0.01 + 0.3 \cdot 0.02 = 1.3\%\end{aligned}$$

□

3.3 Bayes rule

Now consider the example above, and assume that we have got a defect component from the stock (event B). We will derive the probability that the component originates from supplier A_1. We then use Bayes formula that states if $A_1,A_2,\ldots,A_r$ represent a division of the sample space, and B is an arbitrary event then:

$$\Pr(A_j|B) = \frac{\Pr(B|A_j) \times \Pr(A_j)}{\sum_{i=1}^{r} \Pr(A_i) \times \Pr(B \mid A_i)} \tag{7}$$

Example 3.2, continued

We have

$$\Pr(A_1|B) = \frac{\Pr(B|A_1) \times \Pr(A_1)}{\sum_{i=1}^{r} \Pr(A_i) \times \Pr(B \mid A_i)} = \frac{0.01 \times 0.7}{0.013} = 0.54$$

Thus, the probability of A_1 is reduced from 0.7 to 0.54 when we know that the component is defect. The reason for this is that components from supplier A_1 are the best ones, and hence when we know that the component was defect, it is less likely that it was from supplier A_1. □

3.4 Stochastic variables

Stochastic variables are used to describe quantities which we can not be predicted exactly. Note that the word *random quantity* is often used to denote a stochastic variable.

X is stochastic $\Leftrightarrow$ Impossible to predict the value of X

To be more precise, define

- S = Sample space of a random experiment
- e_1, e_2, e_3 are the events comprising the sample space, S = {e_1,e_2, . . }

A stochastic variable X is a real valued function assigning a quantitative measure to each event e_i in the sample space. i.e. $X = X(e_i)$

The function $X = X(e_i)$ is illustrated in Figure 8:

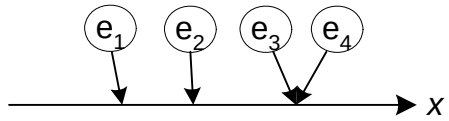


Figure 8 Illustration of a stochastic variable, $X = X(e_i)$

Often the underlying events, e_i are of little interest. We are only interested in the stochastic variable $\underline{X}$ measured by some means.

We sometimes use the word “random quantity” rather than the technical word “stochastic variable”.

Examples of stochastic variables are given below:

- X = Life time of a component (continuous)
- R = Repair time after a failure (continuous)
- Z = Number of failures in a period of one year (discrete)
- M = Number of derailments netxt year
- N = Number of delayed trains next month
- W = Maintenance cost next year

Note
We differentiate between *continuous* and *discrete* stochastic variables. Continuous stochastic variables can take any value among the real numbers, whereas discrete variables can take only a *finite* (or countable finite) number of values.

Probability distribution function

A stochastic variable X is characterized by it's *probability distribution function*

$$F_X(x) = \Pr(X \leq x) \tag{8}$$

We use subscript X to emphasise the relation to the distribution function of the quantity X. The argument (lowercase x) states which values the random quantity X could take. From the expression we observe that $F_X(x)$ states the probability that the random quantity X is less or

equal than (the numeric value of) x. A typicall distriution function is shown in Figure 9. Notate that the distribution function is strictly increasing, and $0 \leq F_X(x) \leq 1$.

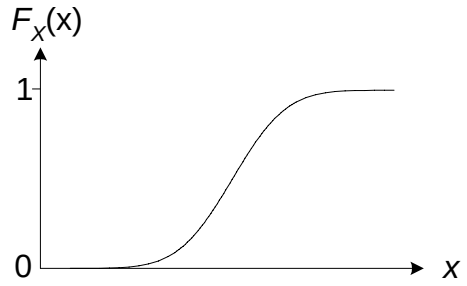


Figure 9 Probability distribution function, $F_X(x)$

From $F_X(x)$ we can obtain the probability that X will be within a specified interval, $[a,b)$:

$$\Pr(a \leq X < b) = F_X(b) - F_X(a) \tag{9}$$

Example 3.3

Assume that the probability distribution function of X is given by $F_X(x) = 1 - e^{-(0.01x)^2}$, and we will find the probability that X is in the interval (100,200]. From Equation (9) we have:

$\Pr(100<X\leq 200) = F_X(200) - F_X(100) = \left[1 - e^{-(0.01\times 200)^2}\right] - \left[1 - e^{-(0.01\times 100)^2}\right] = e^{-1} - e^{-4} = 0.35$ □

Probability density function

For a continuous stochastic variable, the *probability density function* is given by

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) \tag{10}$$

The probability density function express how likely the various x-values are. Note that for continuous random variables the probability that X will take a specific value vanishes. However, the probability that X will fall into a small area around a specific value is positive. For each x-value given in Figure 10, $f_X(x)$ could be interpreted as the probability that X will fall within a small interval around x. Specially we have:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\mathrm{d}u \tag{11}$$

and

$$\Pr(a < X \leq b) = \int_{a}^{b} f_X(x)\mathrm{d}x \tag{12}$$

The last expression is illustrated in Figure 11.

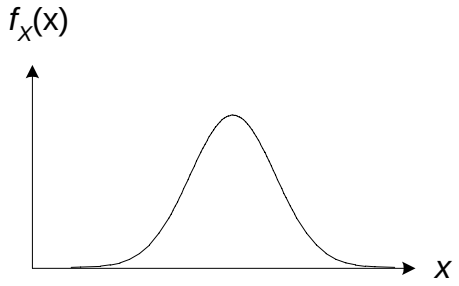


Figure 10 Probability density function $f_X(x)$

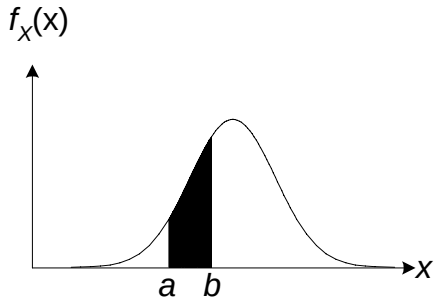


Figure 11 Illustration of Pr($a < X \leq b$)

Random quantities that take discrete values are said to be discretely distributed. For such quantities we introduce the point probability for X in the point x_j:

$$p(x_j) = \Pr(X = x_j) \tag{13}$$

where $x_1, x_2, \ldots$ are possible values X could take.

Expectation

The expectation (mean) of X is given by

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\mathrm{d}x \quad \text{if } X \text{ is continuous}$$

$$E(X) = \sum_j x_j \cdot p(x_j) \quad \text{if } X \text{ is discrete} \tag{14}$$

The expectation can be interpreted as the long time run average of X, if an infinite amount of observations are available.

Variance

The variance of a random quantity express the variation in the value X will take in the long run. We denote the variance of X by:

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 \cdot f_X(x)\mathrm{d}x \quad \text{if x is continuous}$$

$$\mathrm{Var}(X) = \sum_j \left[(x_j - E(X) \right]^2 \cdot p(x_j) \quad \text{if x is discrete} \tag{15}$$

Standard deviation

The standard deviation of X is given by

$$\mathrm{SD}(X) = + \sqrt{\mathrm{Var}(X)} \tag{16}$$

The standard deviation defines an interval which observations are likely to fall into, i.e. if 100 observations are available, we expect that approximate[3] 67 of these observations fall in the interval

$$[E(X)\text{-SD}(X), E(X)\text{+SD}(X)]$$

Precision

The precision (P) is the reciprocate of the variance (V), i.e. $P = \frac{1}{V}$.

α-percentiles

The upper α-percentile, x_α, in a distribution $F_X(x)$ is the value satisfying $\alpha = \Pr(X > x_\alpha) = 1\text{-}F_X(x_\alpha)$.

[3] This result is valid for the normal distribution. For other distributions there may be deviation from this result.

4. COMMON PROBABILITY DISTRIBUTIONS

4.1 The Normal distribution (Gaussian distribution)

X is said to be normally distributed if the probability density function of X is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{17}$$

where μ and σ are parameters that characterise the distribution . It can be shown that:

$E(X) = \mu$

$\text{Var}(X) = \sigma^2$ (18)

The distribution function for X could not be written on closed from. Numerical methods are required to find $F_X(x)$. It is convenient to introduce a standardised normal distribution for this purpose. We say that U is standrad normal distributed if it's probability density function is given by:

$$f_U(u) = \phi(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}} \tag{19}$$

We then have

$$F_U(u) = \Phi(u) = \int_{-\infty}^{u}\phi(t)\mathrm{d}t = \int_{-\infty}^{u}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t \tag{20}$$

and we observe that the distribution function of U do not contain any parameters. We therefore only need one look-up table or function representing $\Phi(u)$. A table is given in the appendix of this compendium.

To calculate probabilities in the non-standardised normal distribution we use the following result:

If X is normally distributed with parameters μ and σ, then

$$U = \frac{X-\mu}{\sigma} \tag{21}$$

is standard normally distributed.

Example 4.1

Let X be normally distributed with parameters $\mu = 5$ and $\sigma = 3$. Find $P(3 < X \le 6)$. We have:

$$\Pr(3 < X \le 6) = \Pr\left(\frac{3-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{6-\mu}{\sigma}\right) = \Pr\left(\frac{3-5}{3} < U \le \frac{6-5}{3}\right) = \Phi\left(\frac{1}{3}\right) - \Phi\left(\frac{-2}{3}\right) =$$

$\Phi(0.33)-(1-\Phi(0.67)) = 0.629-1+0.749 = 0.378$ □

Exercise 2
Let X be the height of men in a population, and assume X is normally distributed with parameters $\mu = 181$ and $\sigma = 4$. How large percentage of the population is more than 190 cm?

4.2 The exponential distribution

For the exponential distribution we have:

$$f_X(x) = \lambda e^{-\lambda x}$$

$$F_X(x) = 1-e^{-\lambda x}$$

$$E(X) = 1/\lambda$$

$$\mathrm{Var}(X) = 1/\lambda^2 \tag{22}$$

Note that for the exponential distribution, X will always be greater than 0. The parameter λ is often denoted the intensity in the distribution

Example 4.2

We will obtain the probability that X is greater than it's expected value. We then have:

$$\Pr(X > E(X)) = 1- \Pr(X \leq E(X)) = 1- F_X(\mathrm{E}(X)) = e^{-\lambda E(X)} = e^{-1} \approx 0.37$$ □

4.3 The Weibull distribution

For the Weibull distribution we have:

$$f_X(x) = \alpha\lambda(\lambda x)^{\alpha-1} e^{-(\lambda x)^{\alpha}}$$

$$F_X(x) = 1- e^{-(\lambda x)^{\alpha}}$$

$$E(X) = \frac{1}{\lambda}\Gamma(\tfrac{1}{\alpha}+1)$$

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}\left(\Gamma(\tfrac{2}{\alpha}+1) - \Gamma^2(\tfrac{1}{\alpha}+1)\right) \tag{23}$$

Where $\Gamma(\cdot)$ is the gamma function. Note that in the Weibull distribution X will also always be positive.

4.4 The gamma distribution

For the gamma distribution we have:

$$f_X(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}(x)^{\alpha-1} e^{-\lambda x}$$

$$E(X) = \alpha/\lambda$$

$$\mathrm{Var}(X) = \alpha/\lambda^2 \tag{24}$$

If we know the expectation, E and the variance, V, of a gamma distribution we could obtain the parameters α and λ by: $\lambda = E/V$, and $\alpha = \lambda \times E$.

4.5 The inverted Gamma distribution

For the inverted gamma distribution we have:

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\left(\frac{1}{x}\right)^{\alpha+1} e^{-\lambda/x}$$

$$E(X) = \lambda/(\alpha\text{-}1)$$

$$\mathrm{Var}(X) = \lambda^2(\alpha\text{-}1)^{-2}(\alpha\text{-}2)^{-1} \tag{25}$$

Note that if X is gamma distributed with parameters α and λ, then $Y = X^{-1}$ has an inverted gamma distribution with parameters α and $1/\lambda$. If we know the expectation, E and the variance, V, of an inveted gamma distribution we could obtain α and λ by $\alpha = E^2/V + 2$, and $\lambda = E(\alpha\text{-}1)$.

4.6 The lognormal distribution

A random variable X is said to have a lognormal distribution if its probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{\tau}\frac{1}{x}e^{-\frac{1}{2\tau^2}(\log x - \nu)^2} \tag{26}$$

We write $X \sim \mathrm{LN}(\nu,\tau)$. The mean and variance of X is given by

$$E(X) = e^{\nu+\frac{1}{2}\tau^2}$$

$$\mathrm{Var}(X) = e^{2\nu}(e^{2\tau^2} - e^{\tau^2}) \tag{27}$$

The following theorem is given without any proof:

Theorem

If X is lognormally distributed with parameters ν and τ, then $Y = \ln X$ is normally distributed[4] with mean ν and variance τ^2.

[4] $\ln(\cdot)$ is the natural logarithm function

4.7 The binomial distribution

Before the binomial distribution is defined, binomial trials are defined.

Let A be an event, and assume that the following holds:

i) n trials are performed, and in each trial we record whether A has occurred or not.
ii) The trials are stochastic *independent* of each other
iii) For each trial Pr(A) = p

When i)-iii) is satisfied, we say that we have binomial trials.

Now let X be the number of times event A occurs. X is then a stochastic variable with a binomial distribution. This is written $X \sim Bin(n, p)$

The probability function is given by

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad for\ x = 1,2,..,n \tag{28}$$

The probability distribution function $\Pr(X \leq x)$ is given in statistical tables.

For the binomial distribution, expectation and variance are given by:

$$E(X) = np$$

$$\mathrm{Var}(X) = np(1-p) \tag{29}$$

4.8 The Poisson distribution

The Poisson distribution is often appropriate in the situation where the random quantity can take the values 0,1,2,.... For the Poisson distribution we have:

$$p(x) = \Pr(X = x) = \frac{\lambda}{x!} e^{-\lambda}$$

$$E(X) = \lambda$$

$$\mathrm{Var}(X) = \lambda \tag{30}$$

It can be proved that the Poisson distribution is appropriate if the following situation applies: Consider the occurrence of a certain event (e.g. a component failure) in an interval (a,b), and assume the following:

i) A could occur anywhere in (a,b), and the probability that A occurs in $(t,t+\Delta t)$ is approximately equal to $\lambda \Delta t$, and is independent of t (Δt should be small).
ii) The probability that A occurs several times in $(t,t+\Delta t)$ is approximately 0 for small values of Δt.
iii) Let I_1 og I_2 be disjoint intervals in (a,b). The event {A occurs within I_1} is then independent of the event {A occurs in I_2}.

When the criteria above are fulfilled we say we have a Poisson point process with intensity λ. The number of occurrences (X) of A in (a,b) is then Poisson distributed with parameter $\lambda(b-a)$, i.e.

$$p(x) = \Pr(X = x) = \frac{\lambda(b-a)}{x!} e^{-\lambda(b-a)} \tag{31}$$

It may also be proven that the times between occurrence of A in a Poisson point process are exponentially distributed with parameter λ.

4.9 The Inverse-Gauss distribution

The Inverse-Gauss distribution is often used when we have an ”under laying” deterioration process. If this deterioration process follows a Wiener process with drift η and diffusion constant δ^2, the time T, until the first time the process reaches the value ω will be Inverse-Gauss distributed with parameters $\mu = \omega/\eta$, and $\lambda = \omega^2/\delta^2$. A Wiener process is shown in Figure 12.

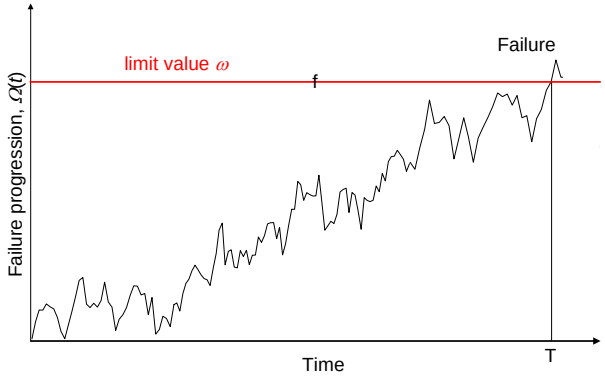


Figure 12 Wiener process

If the failure progression $\Omega(t)$ follows a Wiener process it could be proven that $\Omega(t)$ - $\Omega(s)$ is normally distributed with mean $\eta(t - s)$ and variance $\delta^2(t - s)$. That is η is the average growth rate in the curve, whereas δ^2 is an expression for the variation around the average value.

For the Inverse-Gauss distribution we have:

$$F_T(t) = \Phi\left(\frac{\lambda}{\mu}\sqrt{t} - \sqrt{\lambda}\frac{1}{\sqrt{t}}\right) + \Phi\left(-\frac{\lambda}{\mu}\sqrt{t} - \sqrt{\lambda}\frac{1}{\sqrt{t}}\right) e^{2\lambda/\mu} \tag{32}$$

and

$$E(T) = \text{MTTF} = \mu$$
$$\text{Var}(T) = \mu^3/\lambda \tag{33}$$

5. FAILURES AND FAULT CLASSIFICATION

5.1 Failure

In order to define the term ‘failure’, we need first to introduce the term ‘function’. A unit or system (entity) is designed for performing one or more functions. For example a turnout should be able to direct a train straight forward, or to a deflecting section. A failure is then defined as the event that the possibility of performing the required function is terminated (BS 4778).

5.2 Fault

We use the term ‘fault’ or ’fault state’ to denote the state that the entity is not able to perform its required function.

5.3 Failure mode

A failure mode is defines as the effect a failure has in the way it is observed on the entity that has failed. (EuReDatA).

In some presentations the term (technical) failure mode is used on what we later will denote failure case. This is the case for many RCM (Reliability Centred Maintenance) presentations.

5.4 Failure classification

There are many principle to choose among for classifying failures. In this section we will consider the following dimensions:

- Immediate↔ gradual failure
- Hidden ↔ evident failure
- Physical ↔ Functional failures

5.4.1 Immediate↔ gradual failure

We use the term ’immediate failure’ when the failure occurs spontaneously without any alert. This failure type is often related to situation where the entity is a binary function (only two states). A gradual failure is on the other side characterised by a gradual weakening of the performance, and we are able to observe this weakening.

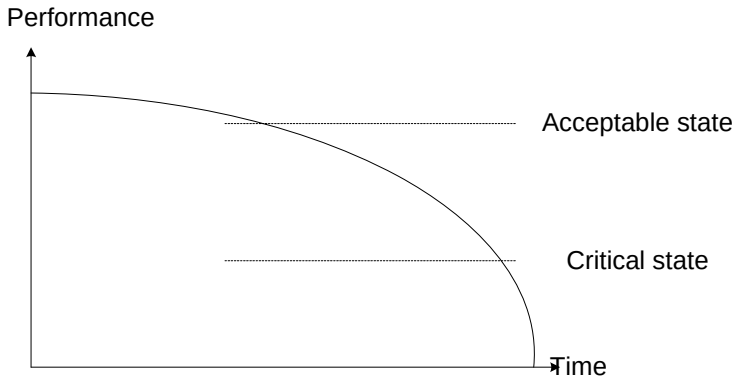


Figure 13Gradually weakening of performance

For safety critical functions ”acceptable state” is often something that is defined based on a assumption of “safe enough”. In this situation we may define failure as the state that the performance not any more is acceptable. Time To Failure (TTF) could in this situation denote the time interval from the entity is put into service until performance no loner is acceptable. However, a “critical situation” is more difficult to define. What is critical, does not only depend on the component performance, but also on the environment. For example if a acceptance limit is defined related to rail wear, this not necessarily mean that if the acceptance limit is exceeded we will have a derailment. The critical situation (derailment) also depends on the wheel profile, the speed of the train, whether it is in a curvature or not and so on.

System analysis such as fault tree analysis, reliability block diagram analysis etc is more complicated when we are dealing with gradually deterioration of components. On component level, a precise fault state is not defined since it depends on the load. Figure 14 shows the situation where the performance (Power of resistance) is gradually weakened. A acceptance level is defined, and overhaul/replacement should be conducted at time T_1. At this point of time it is a very small probability that the load is too high, and thus less risk of derailment. The risk is acceptable, However, if no maintenance is conducted at time T_1, the load will exceed the power of resistance with an increasing probability. At time T_2 it is a significant risk of derailment. We could say that the performance has reached a critical value at time T_2, but an accident will only occur if we experience a load that exceeds the power of resistance.

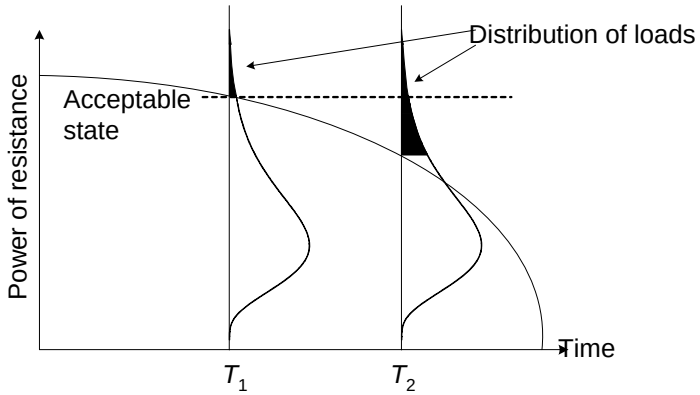


Figure 14 Performance (Power of resistance) in relation to the load

Exercise 3
List 5 examples of immediate failures, and 5 examples of gradually failures related to industrial systems that you are familiar with. □

5.4.2 Hidden ↔ evident failure

We often distinguish between hidden and evident failures. The term 'hidden' often relates to entities that is not continuously demanded. For example the SIFA valve on a train (bleed of the air pressure by activation) is a hidden function, and a failure will not be detected automatically. The term 'evident' relates to entities that are continuously demanded, and a failure will most likely be detected immediately. Note that the same SIFA-valve will also have a evident function ("not bleed of air pressure under normal operation) because an unintended activation immediately will be detected (breaks are activated).

Exercise 4
List 3 examples of hidden failures, and 3 examples of evident failures related to industrial systems that you are familiar with. □

5.4.3 Physical ↔ Functional failures

We also distinguish between physical and functional failures. Physical failures could be eliminated by a repair activity, or by replacing a unit with a new one. Typical causes behind physical failures could be natural ageing (inside the design boundaries), and external load (often outside the design boundaries). A functional failure relates to wrong design, wrong location, wrong usage etc. A replacement with the component with a similar new one will not help. For example if a smoke detector is mounted in an area where there will be no smoke in case of a fire, it will not cure the situation with a new detector at the same location.

5.5 Failure mechanisms and failure causes

Failure mechanisms relates to physical, chemical or other processes that deteriorates the entity, and leads to a failure. The term 'failure cause' is often used in two different ways:

- Failure on a lower level in the system hierarchy, e.g. a defect bearing in a pump
- Root cause, for example bad maintenance, inadequate design etc

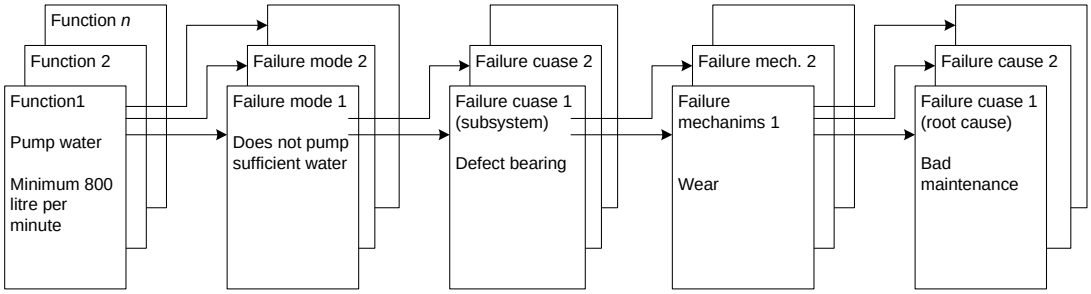


Figure 15 Hierarchy of function, failure mode, failure cause and failure mechanism

In Figure 15 we have visualised the relation between function, failure mode, failure cause (subsystem), failure mechanism and failure cause(root cause). In principle it is a "one to many" relation from left to right.

Exercise 5
Construct a similar illustration as in Figure 15 for the breaking system of a train. Only sketch one function, one failure mode, one failure cause etc. □

5.6 Failure models

5.7 Component reliability

Many methods that are used within RAMS methods are based on the assumption that each component has a binary representation. Such a binary representation express that the component is able to perform the required function, or it fails in performing its function[5]. The state of the component could the be described by a sate variable $x(t)$, where

$$x(t) = \begin{cases} 1 \text{ if the component is functioning at time } t \\ 0 \text{ if the component is in a fault state at time } t \end{cases} \quad (34)$$

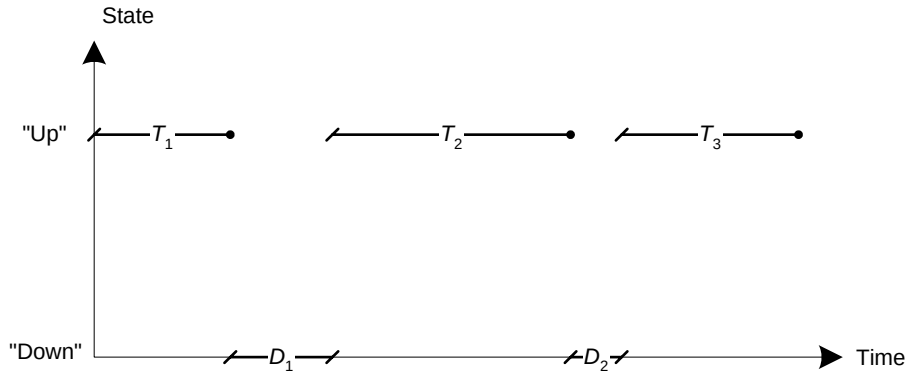


Figure 16 Sate of a component

Figure 16 shows a typical realisation of the sate of a component as a function of the time t. Here the ”Uptimes” are denoted by T_1, T_2 and T_3, whereas the ”Downtimes” are denoted by D_1 and D_2.

5.8 Time to failure (TTF)

The term ‘time to failure’ (TTF) denotes the time from a unit is put into service, until it fails. That is TTF is equivalent to T_1 in Figure 16. In some situations we also use the term time to failure to denote T_2, T_3 etc in Figure 16. It should however be denoted that the distribution of subsequent uptimes are not necessarily identical. The time to failure for a component will be a random quantity (stochastically variable), and we often use the letter T to denote the time to failure. Note that we later will introduce the term service life (SL) to denote the life length of a component regardless of the number of failures. However, for the time to failure, T we could define the following quantities of interest:

- Distribution function, $F(t) = \Pr(T \leq t)$
- Survivor function $R(t) = 1 - F(t) = \Pr(T > t)$

[5] Note that a component could have several functions, and several failure modes. These functions and failure modes has to be identified. In this presentation we assume one only function and only one failure mode.

- Hazard rate $z(t)$
- Mean Time To Failure (without maintenance), MTTF

The distribution function is a function that express the probability that the time to failure, T, is less than or equal to t, i.e. the component fails before or at time t. The survivor function is the probability that the component survive time t t, i.e. the time to failure T is greater than t.

5.8.1 Hazard rate, $z(t)$

To interpret the hazard rate we could use the following relation:

$$z(t)\cdot\Delta t \approx \Pr(t < T \le t + \Delta t \mid T > \mathrm{t}) \tag{35}$$

i.e. the probability of a failure in $(t,t+\Delta t]$ given that the component has survived up to time t.

5.8.2 MTTF and MTBF

Mean time to failure, MTTF, express the time from a new component is put into service until it fails in average. MTTF is only defined if we are talking about the first failure time in Figure 16, or if the subsequent failure times are identically distributed as the first one. If we consider Figure 16 we realise that MTBF = (Mean Time Between Failures) = MTTF + MDT.

Exercise 6

Consider a component where time to failure is exponentially distributed with MTTF = 10 000 hours.

a) What is the probability that the component survive MTTF?
b) What is the probability that a component that has survived MTTF, will survive another 10 000 hours (hint: Pr(A|B) = Pr(A and B) / Pr(B)
c) What is the probability that a component that has survived MTTF will fail within the next hour. □

Exercise 7

a) Repeat exercise **6**, but assume that the component has a Weibull-distributed time to failure with $\alpha = 2$. Hint: Use the fact that $\Gamma(1/\alpha + 1) = \Gamma(3/2) = 0.88623$
b) Compare with exercise **6**. □

5.9 Component availability

We will consider Figure 16 and try to find the unavailability (or availability). Two situations are considered:

- Evident function
- Hidden function

5.9.1 Evident

An evident function means that the a failure of the component immediately will be detected, i.e. when we go from “Up” to “Down” in Figure 16. To obtain the unavailability, U, we intuitively see from Figure 16 that U could be assessed by:

$$U = \frac{D_1 + D_2 + D_3 + \ldots}{(T_1 + D_1) + (T_2 + D_2) + (T_3 + D_3) + \ldots} \tag{36}$$

Now introduce mean time to failure = MTTF = E(T_1) = E(T_2) = …, and mean down time = MDT = E(D_1) = E(D_2) = …, and we observe that the unavailability is given by:

$$U = \frac{\text{MDT}}{\text{MTTF} + \text{MDT}} \approx \lambda \cdot \text{MDT} = \lambda / \mu \tag{37}$$

where we also have introduced:

λ = 1/MTTF = failure rate (assume constant failure rate)
μ = 1/MDT = repair rate (assume constant repair rate) (38)

The component availability is usually denoted A, and we have that A = 1- U. Note that here the mean time to failure is the mean time to failure with a given preventive maintenance level. If we do not do any preventive maintenance, MTTF corresponds to MTTF_N, but if we maintain preventively we should use the effective MTTF, i.e. MTTF_E.

Exercise 8
Find U for a component that fails once a year, and it is required 10 hour to repair it. □

5.9.2 Hidden function – Periodic testing

A hidden function means that a component failure is not immediately detected. In relation to Figure 16 this means that we do not know when we are going from “Up” to “Down”. Thus, in this situation D_1,D_2,… represent the ”non detected” downtime, and the time of repair. In order to reduce the non-detected downtime, the component is tested (function test) periodically, and time between testing is equal to τ.

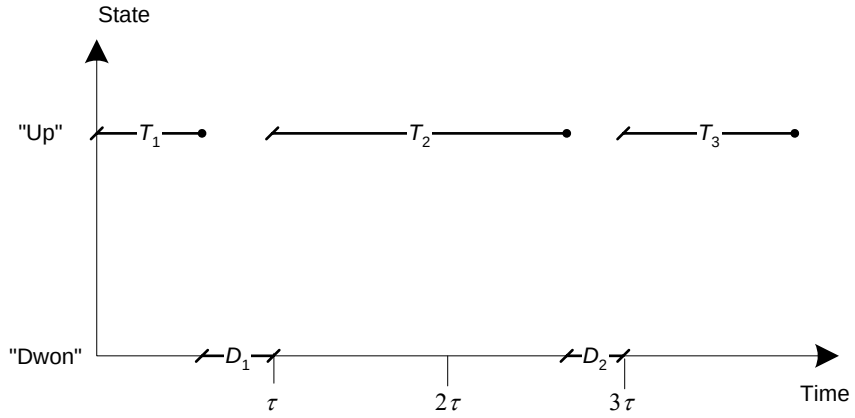


Figure 17 Function test with interval length τ

If the component fails in a period, it will in average have been down half of the interval, i.e. τ/2. If repair time further is short compared to τ, the average downtime MDT is τ/2. Thus, we have:

$$U = \frac{\text{MDT}}{\text{MTTF} + \text{MDT}} = \frac{\tau / 2}{\text{MTTF} + \tau / 2} \approx \frac{\lambda \cdot \tau}{2} = \text{PFD} \tag{39}$$

Where PFD = Probability of failure on demand. Here we have given an intuitive argument for equation (39). To derive the PFD in a general situation we refer to e.g. Rausand and Høyland (2004) where it is shown that:

$$\text{PFD} = 1 - \frac{1}{\tau}\int_0^{\tau} R(t)\mathrm{d}t \tag{40}$$

This result could be used to find PFD for system of several components. For example for a parallel structure of n components we have (Rausand and Høyland 2004):

$$\text{PFD} = \frac{(\lambda\tau)^n}{n+1} \tag{41}$$

if the components are stochastically independent, and each component has a constant failure rate λ, and the test interval equals τ. Note that we usually assume constant hazard (failure) rate. If we have an increasing hazard rate we should replace λ with the effective failure rate, i.e. λ_E if we preventively replace the component (in addition to doing functional tests).

Exercise 9

A unit with hidden function has a constant failure rate $\lambda = 0.3$ per year. How often do we need to test this component in order to fulfil PFD < 1%? □

Exercise 10

Use equation (40) to prove that PFD ≈ $\lambda\tau/2$ in the situation with exponentially distributed time to failure. Hint: $e^{-x} \approx 1 - x + x^2/2$ for small values of x. □

6. LIFE TIME MODELLING

In reliability analysis we are often interested in *life times* of a component or a system. Life times can be treated as stochastic variables. Life times are restricted to non negative values, and are thus a more narrower class than stochastic variables. In this chapter we list basic definitions related to life times:

Function: The function of a system or component is the main task the system/component is designed for. A system may have several functions. For example the function of a valve can be to both *regulate* the flow, and to *stop* the flow.

Failure: The termination of a systems ability to perform a required function.

Life time: The concept of *life time* applies only for components which are discarded the first time they fail. The life time of a component is the time from the component is put into service until the component fails. The life time of a component is treated as a random variable. We will in this context use the capital X to represent life times. Parametric models, such as the Weibull and exponential distributions are used to describe the distribution of this random variable.

Life time distribution: The life time distribution of a stochastic variable X is given by $F_X(x) = P(X \leq x)$. The mathematical expression for $F_X(x)$ contains *parameters*. The goal of life time analysis is to estimate these parameters and identify relevant parametric distributions. Examples of life time distributions are the exponential, Weibull, Lognormal and Gamma distributions.

Censored life time: The life time of a component is defined to be the time from the component is put into service until it fails. In many situations we are prevented from observing the full life time. One example would be when the component has not failed at the termination of the experiment. We differ between *left* and *right* censoring. Left censoring means that we do not exactly know when the component was put into operation. Right censoring means that we know that the component has survived up till some time, say T, but we do not know the history after T.

Hazard rate: The hazard rate given by $z_x(x) = \dfrac{f_X(x)}{1 - F_X(x)}$

where $f_X(x) = \dfrac{dF_X(x)}{dx}$ is the probability density function of X, and $F_X(x)$ is the distribution function of X. Here X is a *life time,* and we use the letter X to indicate that X always should be relatively to the *local* time. With local time we either mean the time for the *first* failure, or time elapsed since the last failure. To best interpret the hazard rate, write:

$$z_x(x)\Delta x \approx P(x < X \leq x + dx \mid X > x)$$

i.e. $z_X(x)\Delta x$ is the probability that a component which has survived up to time x (from system start-up, or since last failure), fails in the interval $(x, x+\Delta x]$.

In classical life time analysis, the hazard rate is identical to the *failure rate*. Other notation for the hazard rate is Force of Mortality (FOM).

Increasing Hazard Rate (IFR): If the hazard rate is non decreasing, we say we have an IFR distribution. The notation is due to the classical use of the word "failure rate".

Decreasing Hazard Rate (DFR):: If the hazard rate is non increasing, we say we have a DFR distribution.

Bath tub shape of the hazard rate. In Figure 18 a bath tub shaped hazard rate is illustrated. Many components are believed to have a bath tub shaped hazard rate.

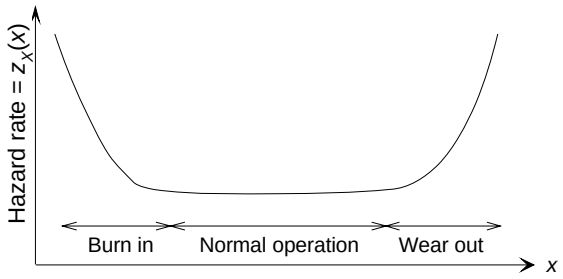


Figure 18 Bath tub shape of the hazard rate

Failure rate: The failure rate is used when the hazard rate, $z_X(x) = \lambda$ =constant. In this situation the symbol λ is used for the failure rate.

Survival probability: The survival probability of a component is the probability that the component survives the time interval from 0 to x, i.e. $R(x) = P(X>x) = 1-F_X(x)$.

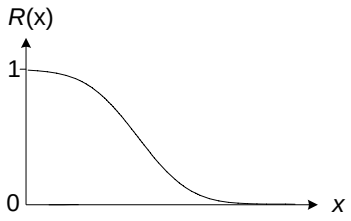


Figure 19 Survival probability, $R(x)$

Mean time to failure (MTTF): The mean time to failure for a component with life time distribution F is defined by:

$$MTTF = E(X) = \mu = \int_0^\infty x\, f_X(x)dx = \int_0^\infty [1 - F_X(x)]dx = \int_0^\infty R(x)dx$$

If the hazard rate, $z_X(x) = \lambda$= constant, i.e. X is exponentially distributed, we have the familiar result:

$$MTTF = \frac{1}{\lambda}.$$

Repair time: The *active* repair time is defined to be the time from a repair action starts until the component is repaired after a failure. The repair time (R) is usually considered as a stochastic variable with some probability distribution, $F_R(\cdot)$.

Mean time to repair (MTTR): The mean time to repair is defined by:

$$MTTR = \int_{t=0}^{\infty} [1 - F_R(t)]dt$$

Down time: The time from a component fails until it is up and running again. The down time includes both waiting time before a repair action starts, the repair time, and time for testing etc.

Mean Down time (MDT): Mean Down Time including, waiting, repair and testing.

Note
Often the words *repair time* and *down time* are used interchangeably. This is not correct unless waiting time can be ignored. For availability calculations the *down time* should be used rather than the *repair time* when waiting time is significant.

Example 6.1 - Exponential distribution

For the exponential distribution we have

$$z(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1-F(t)} = \frac{\lambda e^{-\lambda t}}{1-(1-e^{-\lambda t})} = \lambda \tag{42}$$

This means that the exponential distribution has a constant hazard rate, or FOM. This again means that an old unit is as good as a new unit in terms of statistical performance.

Example 6.2 - Weibull distribution

For the Weibull distribution we have

$$z(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1-F(t)} = \frac{\alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}}}{1-(1-e^{-(\lambda t)^{\alpha}})} = \alpha\lambda(\lambda t)^{\alpha-1} \tag{43}$$

We observe that for $\alpha > 1$ the hazard rate is increasing. For $\alpha < 1$ the hazard rate is decreasing, for $\alpha = 1$ the hazard rate is constant.

7. FAILURE MODELS RELEVANT TO MAINTENANCE

7.1 Introduction

In this section we will present failure models that is relevant to preventive maintenance. Especially these models will be used qualitatively when we identify maintenance action in connection with the RCM logic, but also in relation to optimisation of maintenance intervals.

7.2 The four basic failure models related to preventive maintenance

In this section we will describe four situations that are relevant when modelling life times in relation to maintenance strategies.

7.2.1 Observable gradual failure progression

In this situation we assume that it is possible to observe failure progression prior to the final failure of a component. Consider a pump that is designed to pump 800 litre per minute, and that the pump system is provided with a flow meter. Further assume that it we required a pump capacity of minimum 600 litre per minute to ensure full production. A failure is then defined as the point of time where the capacity of the pump goes below 600 litre per minute. Since we have readings from the flow meter, it is possible to continuous monitor the failure progression. The situation is illustrated in Figure 20.

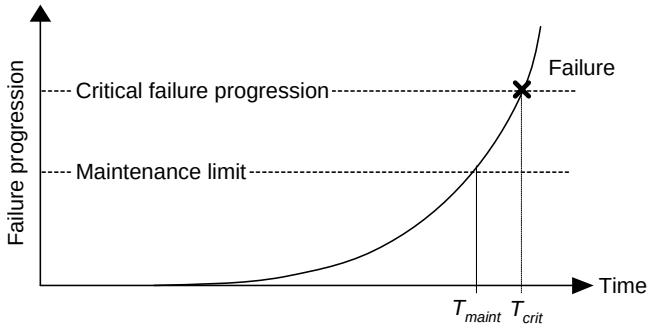


Figure 20 Observable gradual failure progression

To prevent unnecessary failures in Figure 20 we have also illustrated a maintenance limit, where we would replace or overhaul the component. For example when the pump capacity goes below 650 litre per minute we would overhaul the pump. There are two principal questions related to maintenance in this situation:

- What is a reasonable *maintenance limit*?
- How *often* should we monitor or inspect the system?

The more often we inspect and the lower the maintenance limit in Figure 20 is, the lower will the probability of experience a failure be. However, many inspections and a low maintenance limit will imply a very high maintenance cost. We will later develop methods for optimising maintenance in this situation. Note that if no maintenance is carried out, the time to failure will have an increasing failure rate (IFR).

Further note that there might be two types of information about the failure progression:

- Information directly related to the performance of the unit, e.g. the actual capacity of a pump.
- Indirect measures like vibration, temperature increase, particles in the oil etc.

7.2.2 Observable "sudden" failure progression

The situation now is similar to the situation in the previous section, but we now assume that the system could operate for a very long time without any sign of a potential failure, but then at some point of time a potential failure would be evident as illustrated in Figure 21.

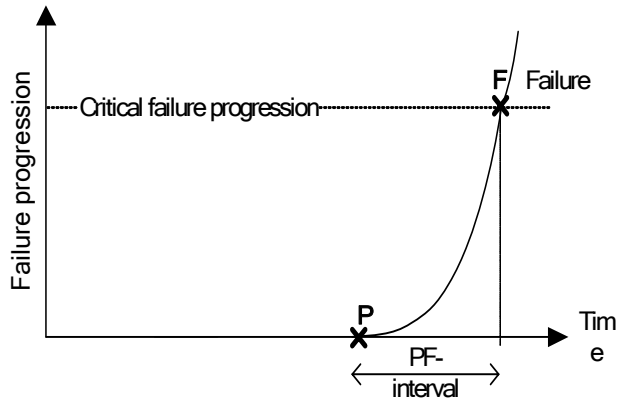


Figure 21 Observable "sudden" failure progression

In Figure 21 we have indicated a "P" for potential failure, i.e. the time where a coming failure is observable. The time interval from the failure is first observable, and till a failure occurs is very often denoted the *PF-interval*. We will in the following denote this situation for the "PF" situation because the PF-interval will be central in the understanding of effective maintenance strategies. An example could be a rail which is exposed to a combination of fatigue and a flat wheel which initiates a crack (potential failure, P). However, such cracks could be detected by ultrasonic inspection, and hopefully we will detect the crack before it propagates to a failure. Note that if no maintenance is carried out, the time to failure will have an increasing failure rate (IFR).

7.2.3 Non-observable failure progression

Assume we have a situation like in section 7.2.1 or 7.2.2, but that we for some reason could not observe the failure progression. For example in the situation with the pump we do not have a flow meter available, or consider a rail with fatigue, but where we are not able to monitor a crack due to no available equipment for ultrasonic inspection. Another situation is wear inside a closed bearing. The situation is illustrated in Figure 22, where we have shown a dashed line for the failure progression due to the fact that it is not observable.

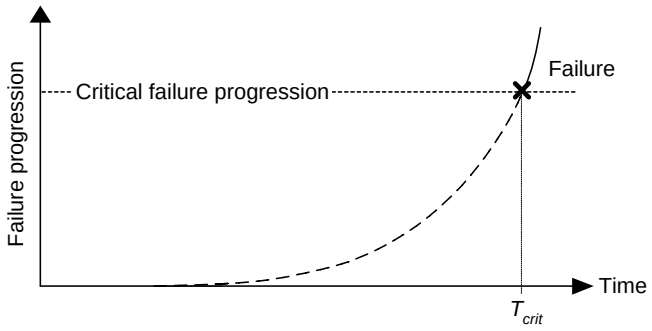


Figure 22 Non-observable failure progression

Since there is ageing phenomenon behind this failure situation, the distribution of the time to failure will have an increasing failure rate. An appropriate maintenance action in this situation would be to replace the component periodically. However, since we are not able to observe failure progression, the time elapsed since the previous maintenance is the only indicator of a coming failure.

7.2.4 Shock

The situation now is similar to the PF-interval situation in Section 7.2.2, but now the PF-interval is extremely short, and there is no possible inspection methods that are able to reveal a potential failure in due time. In this situation, the time to failure will be approximately exponentially distributed.

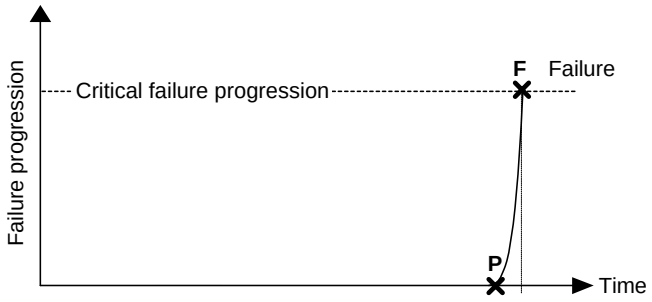


Figure 23 Shock model

7.3 Effective failure rate as a function of maintenance

We will now review the four different situations in Section 6 and investigate the relation between the effective failure rate, and the amount of maintenance carried out.

We will use the following notation:

τ Maintenance interval, either inspection interval, or renewal interval.

$\lambda(\tau)$ Effective failure rate if the component is maintained with interval length τ. We will use subscript to discriminate between different failure models, e.g. $\lambda_{GF}(\tau)$ is used in the situation of gradual observable failure progression.

$Q_I(\tau)$ Probability of not detecting a potential failure in due time if inspected with interval length τ.

PFD Probability of failure on demand, i.e. the average time a failure of a hidden function is not detected. Sometimes the acronym MFDT (Mean Fractional Dead Time) is used rather than PFD.

7.3.1 Observable gradual failure progression

Model for gradual degradation

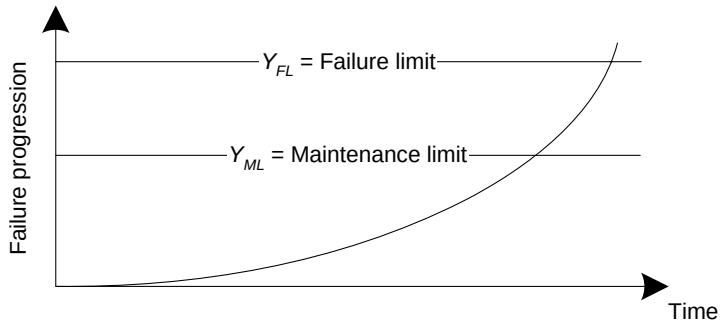


Figure 24 Model for gradual degradation

The failure progression, $Y(t)$, is a random variable. When $Y(t)$ exceeds the maintenance limit (Y_{ML}) a corrective maintenance action is performed which resets the system, i.e. $Y = Y_0$. If $Y(t)$ exceeds the failure limit (Y_{FL}) a failure occurs. $\{Y(t)\}$ could be specified in various ways. Two common used models for $\{Y(t)\}$ are the Wiener process, and the Gamma process. A limitation in these processes is that the degradation is assumed to be linear with time. This is problematic when e.g. cracks are modelled, since the failure progression is believed to go faster and faster as the crack size increases. Another way to specify the degradation is a shock model, see e.g. Aven and Jensen (1999) pp. 79-82 or Rausand and Høyland (1994) p. 246. In a shock model we assume a point process of shocks with some intensity, say ρ, and where the damage at shock i is a random variable, say V_i. In the basic shock models the damage at shock i is independent of the accumulated damage up to shock i which would not be the case for e.g. cracks. Therefore we also for the shock model need to model V_i as a function of $\Sigma_{i-1} V_i$. When modelling the degradation we have two challenges:

Given the parameters in the underlying failure model the aim is to establish a mathematical model that shows the relations between the effective failure rate and the maintenance interval and maintenance limit, i.e. $\lambda = \lambda_{GF}(\tau, Y_{ML})$. Note that there are two approaches for obtaining the parameters describing the underlying failure model:

1. To obtain the underlying parameters we ideally want to observe the process, i.e. the failure progression as a function of time and then establish a failure model with the relevant parameters based on traditional estimation procedures.
2. In real life, we very often have no explicit information about the process in term of failure progression as function of time. However, we could have some assessment of the time it

will take to reach certain levels of degradation. For example in Figure 24 we have mean time and standard deviation for the time it takes to reach Y_{ML} and Y_{FL}. Based on such information we could then "calculate" the required parameters in the failure progression model if we specify the model, i.e. a Wiener process. In the following presentation we assume that we could obtain such data, and that we do not have access to the "real" process parameters.

Now we will return to the modelling of the degradation process. The way we have chosen to specify $\{Y(t)\}$ is as follows:

- The Y-axis is divided into n intervals. The first interval is $I_1 = [Y_0, Y_1)$, the second is $I_2 = [Y_1, Y_2)$ up to the last interval $I_n = [Y_{n-1}, Y_n) = [Y_{n-1}, Y_{FL})$, see Figure 25.
- For each interval I_j we specify the corresponding time, T_j it will take to move from Y_{j-1} to Y_j. T_j is a random quantity which we specify in terms of the expectation E_j and standard deviation SD_j.

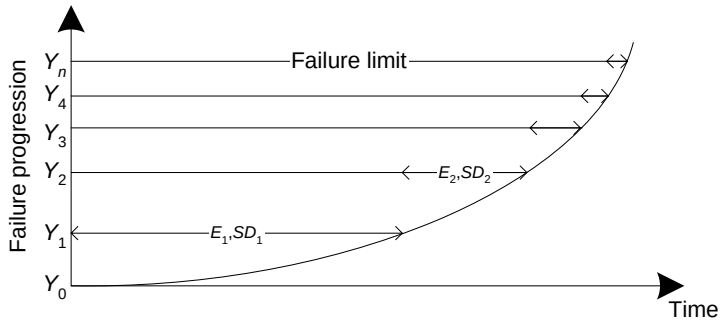


Figure 25 Specification of time to move from Y_{i-1} to Y_i

We will also allow the model to handle "fast failure progressions". With this we mean that for all values of Y, there is a likelihood that a "fast failure progression" starts. We specify this likelihood by a frequency, f_Y. Further, if such a fast failure progression starts, we specify the time until the failure limit, Y_{FL} is reached in terms of expectation and standard deviation of the corresponding PF-interval, see Figure 26. To simplify, we only specify these quantities as average values within each interval of the Y-axis, i.e.

- f_j = frequency of "fast failure progression" in interval I_j
- $E_{PF,j}$ expected PF- interval for the "fast failure progression"
- $SD_{PF,j}$ standard deviation o the PF-interval for the "fast failure progression"

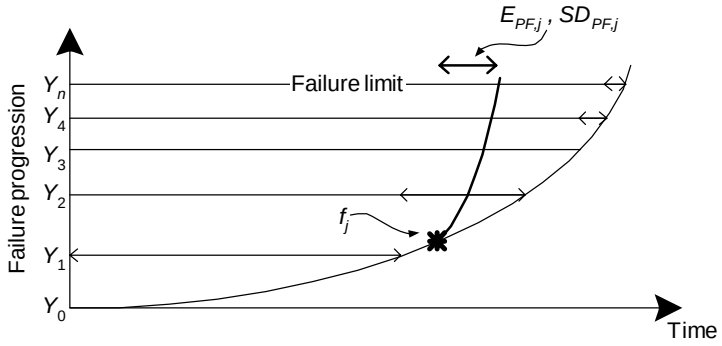


Figure 26 Possibility of "fast failure progression"

The modelling challenge is now to calculate the effective failure rate, λ_{GF} as a function of the inspection interval, τ, and the maintenance limit, $Y_{ML,}$ i.e. $\lambda_{GF} = \lambda_{GF}(\tau, Y_{ML})$.

We might also, in principle, calculate λ_{GF} as a function of a decreasing inspection interval, τ, due to the fact that we will inspect more often as we reach the maintenance limit, Y_{ML}. In the following we present some ideas for solution.

Modelling failure progression as a Wiener process.
We first assume that there is only one interval for which we have specified the mean value and standard deviation. Further assume that "fast failure progression" could not occur. The Wiener process is linked to the inverse Gaussian distribution as follows:

Let $W(t)$ be governed by a Wiener process $\{W(t);\ 0 < t < \infty\}$ with drift η and diffusion constant δ^2. The increment in the Wiener process during a time period Δt is $N(\eta\Delta t, \delta^2\Delta t)$. Further, define the time T until $W(t)$ reaches the value ω for the first time. It could now be shown (Cox and Miller 1965) that T is inverse Gaussian distributed with parameters $\mu = \omega/\eta$ and $\lambda = \omega^2/\delta^2$. For the inverse Gaussian distribution we have:

$$F_T(t) = \Phi\left(\frac{\lambda}{\mu}\sqrt{t} - \sqrt{\lambda}\,\frac{1}{\sqrt{t}}\right) + \Phi\left(-\frac{\lambda}{\mu}\sqrt{t} - \sqrt{\lambda}\,\frac{1}{\sqrt{t}}\right)e^{2\lambda/\mu} \tag{44}$$

and

$$E(T) = \text{MTTF} = \mu$$

$$\text{Var}(T) = \mu^3/\lambda \tag{45}$$

Thus, if we know the mean value (E), and the standard deviation (SD) of T, we could calculate the parameters in the Wiener process by:

$$\eta = \omega/\mu = \omega/E$$

$$\delta^2 = \omega^2/\lambda = \omega^2 \times SD^2/E^3$$

We now want to approximate the Wiener process by an infinite discrete process, Z_t.

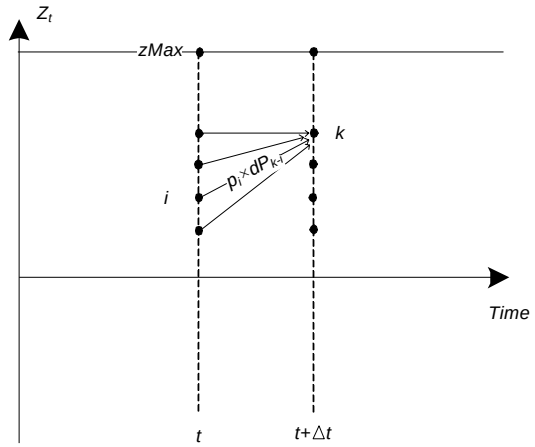


Figure 27 Discrete model: change of state probabilities in an interval of length Δt

We let by definition $Z_t = 0 \Leftrightarrow W(t) = 0$, $Z_t = zMax \Leftrightarrow W(t) = \omega$. Note that Z_t could take infinite low values. Let $p(i,t) = P(Z_t = i)$. Since the increment in the Wiener process during a time period Δt is $N(\eta\Delta t, \delta^2\Delta t)$, we could calculate $p(k,t+\Delta t | Z_t = i)$ by a direct argument. Thus we could find the distribution of an state at any time, $p(i,t)$, by integrating from t=0. We also note that $p(zMax,t)$ is the probability that the process has reached ω for the first time at time t. This means that $p(zMax,t) = P(T<t)$, and we could easily obtain $E(T)$ by integrating $(1-p(zMax,t))$ from 0 to ∞. If the process is not intervened, we know that $E(T) = \mu$ which we could verify numerically. The following elements are required in the calculation procedure:

1. Define the number of discrete states of Z_t, this means define the number $zMax$. Also define a minimum value $zMin$, where $zMin$ is a negative number such that it is very unlikely that Z_t is less than $zMin$ for any value of t. Typically, set $zMax$ = 200, and $zMin$ = - 70.
2. Define the (normalized) step length, i.e. $dz = 1/zMax$
3. Define $JumpMax$ to be the maximum number of steps the process is allowed to jump upwards in an interval of Δt. Similarly define $-JumpMax$ be the number of steps the process is allowed to jump downwards in an interval of Δt. Typically, set $JumpMax$ = 60, and $JumpMin$ = -50.
4. Now we define the steplength, Δt, such that the probability of jumping more than $JumpMax$ steps could be neglected.
5. Calculate the probability of jumping k steps, $k = JumpMin, \ldots, JumpMax$. Since the increment is normally distributed, we have $dP_k = \Phi[\{(k+½)dz - \eta\Delta t\}/\{\delta\sqrt{\Delta t}\}] - \Phi[\{(k-½)dz - \eta\Delta t\}/\{\delta\sqrt{\Delta t}\}]$. dP_k should be stored in an array for later use.
6. Define initial state probabilities: $p_i = 1$ if $i = 0$, else $p_i = 0$. p_i should be stored in an array for continuous updating.
7. Now, the process could be integrated, i.e. we increment t by Δt from 0 to ∞. For $k < zMax$ we have
   $p_k = \Sigma_{i<zMax} p_i \times dP_{k-I}$
   For $k = zMax$ we have
   $p_{zMax} = \Sigma_{i<zMax} p_i \times \Sigma_{j\geq 0} dP_{zMax+j-i}$
8. Now, for each step we let MTTF = MTTF + $(1-p_{zMax})\Delta t$, with MTTF = 0 as an initial value. Thus, we could estimate $E(T)$ = MTTF.

The model could be extended by allowing the system to be maintained when the maintenance limit, Y_{ML} is reached during an inspection. The corresponding discrete value is found to be $zML = zMax \times Y_{ML}/Y_{FL}$

To take this effect into account, we will log the actual time from $t = 0$ during the integration process. If $t \in [j\tau - \frac{1}{2}\Delta t, j\tau + \frac{1}{2}\Delta t]$, $j = 0,1,2\ldots$we know that this time interval represent an inspection. Thus for this interval we define an absorbing state, say *zMaint*, and we let $p_{zMaint} = \Sigma_{zML<i<zMax} p_i$, and thereafter let $p_i = 0$ for $zML < i < zMax$.

As t approaches ∞, p_{zMax} is the probability that the system fails, and p_{zMaint} is the probability that maintenance prevents a failure. To obtain the average cycle length, we could as an approximation use $\mu \times Y_{ML}/Y_{FL}$. We also have the "effective failure rate" $\lambda_{GF} = \lambda_{GF}(\tau, Y_{ML}) = p_{zMax} \times Y_{FL}/(\mu \times Y_{ML})$.

Modelling failure progression as a Wiener process with non constant drift and diffusion parameters

The framework we have set up, will also need to model the situation when the drift and diffusion parameters varies as indicated in Figure 25. The principles of the modelling will be similar to what was done above. However, we now have to calculate a set of dP_k-arrays, one for each interval I_j in Figure 25. The bookkeeping is more complicated, but the principles are the same.

Fast failure progression

We could also model "fast failure progression" in the same way as indicated above. We now have to introduce a parallel state vector representing the situation that we have entered "fast failure progression". For each step, we calculate the probability that we move from "ordinary failure progression" to "fast failure progression". Given that we are in interval I_j, the transition probability (from "ordinary" to "fast") equals $f_j \times \Delta t$. Note also that in the parallel state vector, we need a separate calculation of dP_k that represent "fast failure progression".

Modelling failure progression by a finite state model

In this situation we will assume that the state variable only takes a finite number of states. We first present the model when no maintenance is carried out, i.e., we start at time t=0 and observe the system until failure. We now let:

$$Y(0) = y_0$$

$$Y(T) = y_r \tag{46}$$

where T per definition is the time to failure. Between y_0 and y_r there are r-1 intermediate sates. By choosing a large value of r we could obtain a very good approximation to a continuous process if this is required. We will now let T_i, $i = 0,\ldots,r$-1 define the time the system is in state i. For modelling purposes it will now be very convenient to assume that the T_i's are independent and exponentially distributed with parameter λ_i. In a more general setting we would need to model the T_i's with more advanced distributions than the exponential distribution in order to have a realistic model for the failure progression. However, this will not be pursued in this presentation. We will also assume that the process runs through all states chronologically from y_0 to y_r without "stepping back" at any time.

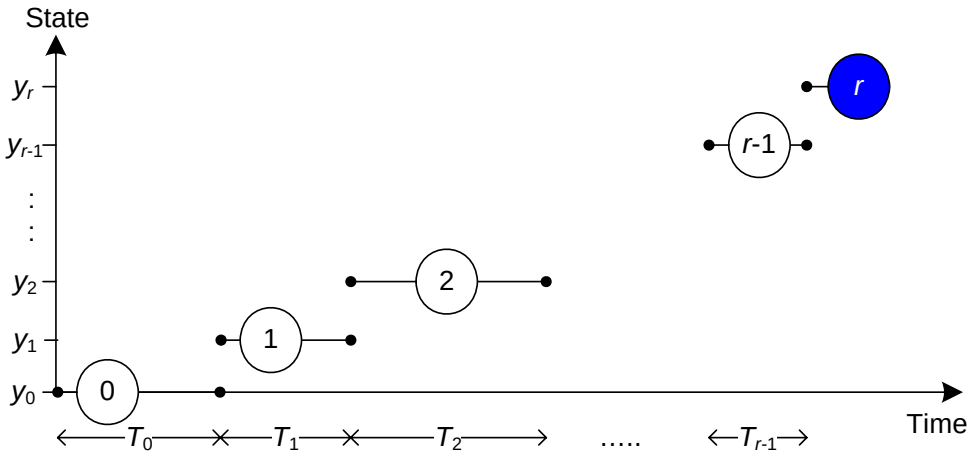


Figure 28 Markov state model

We will now find the probability that the system is in the various states as a function of the time *t.* We let $P_i(t)$ denote the probability that the system is in state *i* at time *t.* By standard Markov considerations (see e.g. Rausand & Høyland 2004), we obtain the Markov differential equations:

$$P_i(t+\Delta t) = P_i(t)(1-\lambda_i \Delta t) + P_{i-1}(t)\lambda_{i-1}\Delta t \quad (47)$$

where Δt is a small time interval, and we set $\lambda_{-1} = \lambda_r = 0$ per definition. Further the initial conditions are given by:

$$P_0(0) = 1$$

$$P_i(0) = 0 \text{ for } i > 0 \quad (48)$$

Equation (47) could easily be integrated by a computer program, for example VBA in MS Excel. See Vatn (2004). It is now easy to find MTTF by another integration, i.e.

$$MTTF = \int_{t=0:\infty} [1-Pr(t)]dt \quad (49)$$

and we should verify that we get $MTTF = \sum_{i=0:r-1} 1/\lambda_i$.

Note that the transition rates (λ_i's) are assumed to be known, that is either they are estimated from data, or found by expert judgment exercises.

RAMS performance as a function of inspection frequency and intervention level
Equation (47) could be used to obtain the MTTF if no maintenance was carried out. In this section we will find the MTTF if *i*) we do a perfect inspection with time between inspection equal to τ, and *ii*) if we intervene at level *l*, i.e. we replace the component with a new component if the state at an inspection is greater or equal than y_l, see Figure 29.

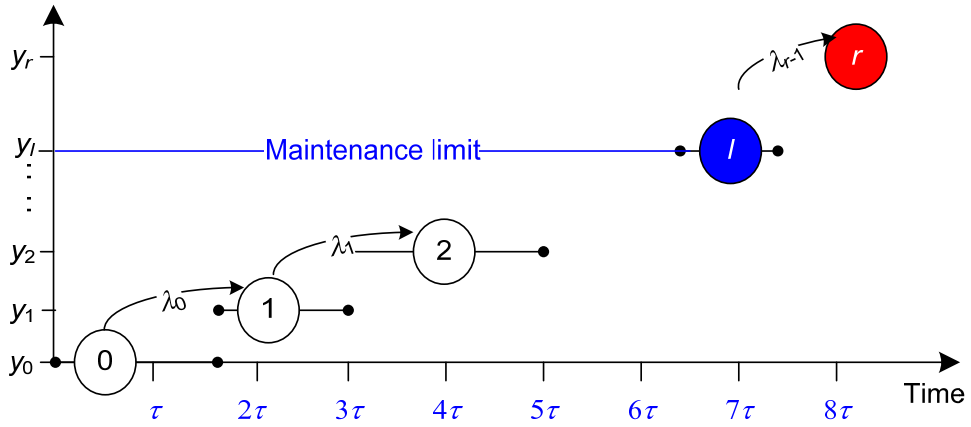


Figure 29 Maintenance limit and inspections in the Markov model

In this situation we also integrate Equation (47), but when t equals τ, 2τ, 3τ,… special considerations are necessary. At these point of times, we will inspect the system, and if $Y(t) \geq y_l$ we will replace the system. This corresponds to setting $P_0 = P_0 + \Sigma_{i \geq l} P_i$, and $P_i(t) = 0$, $i \geq l$. If we now let f be the system failure rate, we use the following procedure to find the MTTF:

1. Define the initial conditions by Equation (48).
2. Set $f = 0$, $t = 0$, Δt = sufficient small.
3. Integrate Equation (47) one step, $t = t + \Delta t$.
4. Let $f = f + P_r(t)$.
5. If $t = \tau, 2\tau, 3\tau, \ldots$, then let $P_0(t) = P_0(t) + \Sigma_{i \geq l} P_i(t)$, and $P_i(t) = 0$, $i \geq l$.
6. Loop to Step 3 until t is sufficient large, i.e. $t >$ "Time horizon".
7. System failure frequency now equals $\lambda_{FP}(\tau,l) = f/t$, and $MTTF = MTTF(\tau,l) = t/f$.

By modifying step 5, we could also calculate the expected number of system replacement (Renewal Rate) by adding $\Sigma_{i \geq l} P_i(t)$ to the expectation at each inspection, and divide the result by the time horizon at the end of the integration.

Modelling imperfect inspection

The procedure in Section 0 assumed that the inspection was perfect, that is an inspection would reveal the correct state with certainty. In practice there will always be a probability that the inspection is not perfect, i.e. that we do not reveal the correct state of the system. Let $Q_{i|j}$ be the probability that the system is classified to be in state i when the real state is j. To obtain the system performance we now only have to replace Step 5 with the following step:

5. If $t = \tau, 2\tau, 3\tau, \ldots$, then let $P_0(t) = P_0(t) + \Sigma_j (\Sigma_{i \geq l} P_j(t)\, Q_{i|j})$.

Modelling an increasing failure intensity

If $Y(t)$ represents a physical measure such as e.g. the length of a crack in a structural element, we know from fracture mechanics that the crack growth speed is increasing with increasing crack size. In this section we will work out some details for how to specify the transition rates in this situation. A simple model that could reflect the increase in crack speed is to let:

$$\lambda_i = \lambda_0 v^i \qquad (50)$$

Where v > 1 determines how much "faster" the crack develops in state i compared to state i-1. Now, let us assume that we want to establish the model in the following situation:

- There are r+1 states, $y_0, y_1, ..., y_r$
- The mean time to failure is known to be $MTTF$
- The variance in the time to failure is known to be Var(T)
- The ratio λ_{r-1}/λ_0 equals V

We first obtain $v = V^{(1/(r-1))}$. Then we use the basic summation rule for a geometric series, and obtain:

$$\lambda_0 = \frac{1 - 1/v^r}{(1 - 1/v)MTTF} \tag{51}$$

in order to fulfil $MTTF = \sum_i 1/\lambda_i$. We could also easily verify that Var(T) is given by:

$$\mathrm{Var}(T) = \frac{1 - 1/v^{2r}}{(1 - 1/v^2)\lambda_0^2} \tag{52}$$

If the calculated variance does not match the variance we are aiming at, we could change the number of states. An increase in the number of states (r) will give smaller variance by Equation (52). To solve the model, we use the same procedure as specified in Section 0.

More on the variability of the failure progression

In the framework we have set up, there are some degrees of freedom when we want to model the state variable. Typically we would like to find parameters such that we achieve the "correct" MTTF and Var(T) for the time to failure when no maintenance is carried out. We may then vary the number of states (r), and the parameter v that describes how much faster the cracks develops at the end compared to the beginning. If the choice of parameters gives a too high variance, we could easily introduce some extra states in the model. However, if the variance found is too low we need to reduce the number of states which could be more problematic since we then could loose the physical understanding of what the states means. For example if the crack length is "forced to" jump in too large steps by the model, we would not believe in the model. If this is the case, we could use the Gamma stochastic process model, or the shock model described in Section **Error! Reference source not found.** which are flexible models allowing us to specify the MTTF and Var(T) according to our need. However, since we want to stick to our framework, we will suggest to model large variability by means of a branch process as described in the following section.

Modelling large variability by means of a branch process

In this section we will describe a branch process that enable us to specify large variance in the time to failure even if we have a large number of states in the model. The idea is that at a certain state the system follows one of several possible branches. Usually this split occurs at time $t = 0$, but we could in principle let the split occur at a later time, or state. Each branch is a Markov chain, where the failure rates describe the transition probabilities given that we follow that branch. A branch could be seen on as a specific type of failure progression, e.g. due to a specific type of crack, a specific location of the crack and so on. Now, assume that there are K different failure progression types, and let:

u_k = Pr("Failure progression type k"), $k = 1,\ldots,K$ (53)

By using the double expectation rule, we find

$$E(T) = MTTF = \sum_k u_k[\sum_i 1/\lambda_{i,k}]$$

$$Var(T) = \sum_k u_k[\sum_i 1/(\lambda_{i,k})^2] + \sum_k u_k (1- u_k)[\sum_i 1/(\lambda_{i,k})]^2 \quad (54)$$

Where $\lambda_{i,k}$ is the transition rate out of state i given that we have a failure progression type k.

We could now identify some "typical" fast failure progression types, and some slow failure progression types, and then choose probabilities for each branch, u_k, in such away that Equation (54) fulfils our need.

The state equations will now be more complicated, but Equation (47) will form a master. We need to replace the $P_i(t)$'s with $P_{i,k}(t)$, and letting $P_{0,k}(t) = u_k$. It is outside the scope of this presentation to give the details here.

Note that this model has two features. Firstly, we are now enabled to specify a large variability in the time to failure which would be necessary in some situations. Secondly, we could also think about the branching process as a physical feature, i.e. we consider each branch as a physical degradation depending on which crack that initially was there. In an advanced inspection strategy, we could also use the information about how fast the identified crack develops to change the inspection interval adaptively. For example, if we see a crack that is developing very fast, we could increase the inspection frequency, and maybe also change the intervention level.

7.3.2 Observable "sudden" failure progression

The situation is as in Figure 21 in Section 7.2.2. The point "P" is the first point in time where we are able to reveal the outset of a failure. When the progression is above some value a breakage/failure will occur (point "F"). The length from a potential failure is detectable until a failure occurs is denoted the PF-interval T_{PF}. The length of the PF-interval is assumed to vary from time to time. For example, consider a rail where a crack can be initialized at different places of the rail, and thus time before the crack "reaches the surface" will vary. Another situation is where the crack propagation depends on the load, e.g. number of heavy axels. The length of the PF-interval should therefore be treated as a random quantity as illustrated in Figure 30.

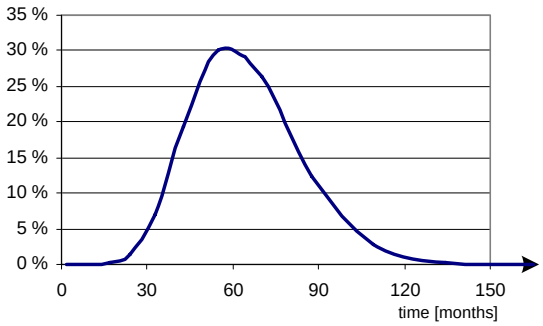


Figure 30 Variation in the PF-interval

Periodic ultrasonic inspection is conducted at intervals of length τ to detect potential failures. The length of the inspection intervals should not be longer than the average PF-interval. However, since the PF-interval varies from time to time, and because there is also probability that a potential failure is not revealed during an inspection, the inspection interval should be shorter than the average PF-interval. A prerequisite for using the PF-intervals in maintenance planning is that a failure is alerted by some degradation in performance, or some indicator variable is alerting about the failure. Such a variable could be vibration, cracks, increased temperature etc.

The following quantities will be relevant when calculating the effective failure rate as a function of the maintenance interval:

- Mean PF-interval length, E_{PF}
- Standard deviation in PF-interval length, SD_{PF}
- Probability that an existing crack (or another warning situation) will be detected by a inspection, P_I (given that it is possible to detect the crack by condition monitoring method)
- Coverage of the inspection method, i.e. percentage of cracks that could be detected, P_C
- Interval length of inspections, τ
- Frequency of potential failures, f_P

Note that P_I could be treated as in ”independent” detection probability, i.e. we have this probability every time vi perform an inspection. P_C is the portion of cracks that could be detected, i.e. coverage of sensors, which types of cracks that could be seen etc.

The effective failure rate could now be found by:

$$\lambda_{PF}(\tau) = f \times Q_I(\tau) \tag{55}$$

where $Q_I(\tau)$ could be interpreted as a “barrier probability”, i.e. the probability that inspection is an efficient barrier. $Q_I(\tau)$ comprises an “independent” term related to each (independent) inspection, and an “dependent” term representing lack of coverage of the inspection method, i.e.

$$Q_0(\tau) = Q_{PF}(\tau, E_{\text{PF}}, SD_{\text{PF}}, P_I) + (1- P_C) \tag{56}$$

Where again Q_{PF} (τ; E_{PF}, SD_{PF}, P_I) could be found by using the provided Excel sheet (PFCalc.xls), or reading Figure 31. See APPENDIX A for a procedure for calculating $Q_0(\tau)$.

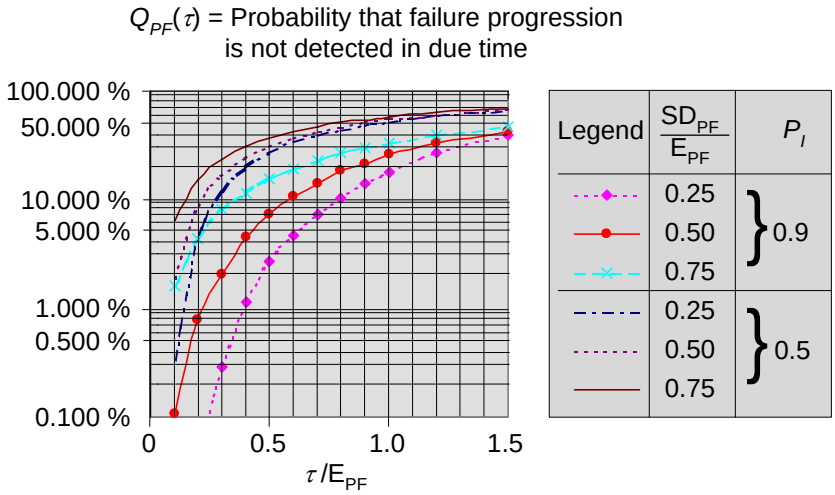


Figure 31 $Q_{PF}(\tau)$ for different combination of SD_{PF}/E_{PF} and P_I

Example 7.1 - Calculation of $Q_{PF}(\tau)$

On a section of a line in average f = 3.5 cracks are initiated every year. The coverage of the detection method is P_C = 0.9, and the (independent) detection probability is P_I = 0.9. Mean and standard deviation of the PF-interval is E_{PF} = 5 years and SD_{PF} = 3 years. Ultrasonic measurement is performed once a year (τ = 1 year).

We now have SD_{PF}/E_{PF} = 0.67, and τ/E_{PF} = 0.2. Further from Figure 31 we now read $Q(\tau = 1; E_{PF} = 5, SD_{PF} = 3, P_I = 0.9)$ = 2%. Finally we obtain $\lambda_{PF}(\tau) = f \times [Q_{PF}(\tau = 1; E_{PF} = 5, SD_{PF} = 3, P_I = 0.9) + (1- P_C)]$ = 3.5 × (0.02 + 0.1) = 0.49, i.e. we would expect a rail breakage every second year. □

7.3.3 Non-observable failure progression

This situation is related to ageing, but we are not able to observe the ageing. To model the effective failure rate in this situation we start with the hazard rate, $z(t)$. We assume that the hazard rate is increasing (IFR), and we measure the ageing in terms of the ageing parameter α. No ageing corresponds to a constant hazard rate (ageing parameter = α = 1). Low (α = 2) and moderate (α = 3) ageing corresponds to a rather undefined point where the hazard rate really starts to increase. Strong ageing (α = 4) means a significant increase of the hazard rate after some time. The four situations are illustrated in Figure 32.

Safe Time to Failure

Sometimes the term 'safe time to failure' (STTF) is introduced. In fact it would be very difficult to interpret such a term, but the term is often used in RCM analysis to denote a point of time where it is very likely that the component survives this point of time. We often define the STTF as the point of time where it is at most 1% probability that the unit fails prior to this time.

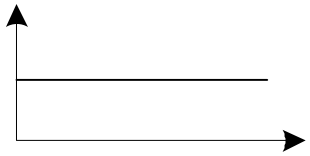
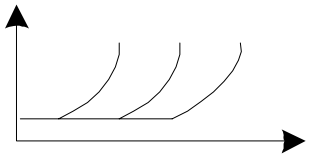
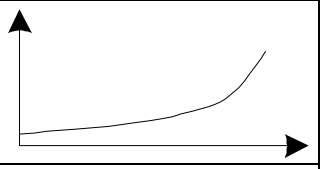
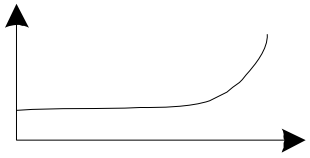
Degree of ageing	Ageing parameter	Graphical situation
No aging	$\alpha = 1$	
Weak aging	$\alpha = 2$	
Medium aging	$\alpha = 3$	
Strong aging	$\alpha = 4$	

Figure 32 Different degrees of ageing

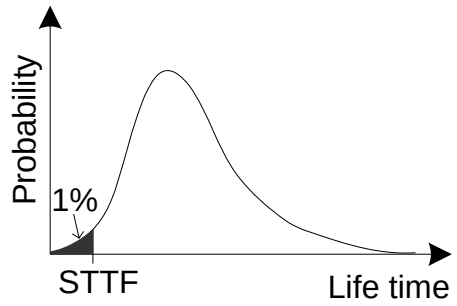


Figure 33 Safe Time To Failure

If the term STTF is used in an RCM procedure this is often an alternative to specify the ageing parameter explicitly. If the ratio MTTF/(MTTF-STTF) is large the ageing is strong, whereas if this ratio is small the ageing is low.

The functional relation between the ageing parameter α and the ratio ρ = MTTF/(MTTF-STTF) is given by:

$$\alpha \approx -1.4\rho^2 + 9.2\rho - 6{,}6 \qquad (57)$$

This formula could be used for "exact" calculation. If it is sufficient with an approximate expression for the ageing parameter Table 1 could be used.

Table 1 Relation between MTTF, STTF and the ageing parameter

MTTF/(MTTF-STTF)	Ageing parameter α	Ageing
< 1.2	2	Low
1.2 – 1.4	3	Moderate
> 1.4	4	Strong

Now, if we know the mean time to failure, MTTF (without maintenance), and the ageing parameter (α) the effective failure rate could be approximated by:

$$\lambda_A(\tau) = \left(\frac{\Gamma(1+1/\alpha)}{\text{MTTF}} \right)^{\alpha} \tau^{\alpha-1} \tag{58}$$

where $\Gamma(\cdot)$ is the gamma function. The approximation is good when the maintenance interval is small compared to the MTTF. If the maintenance interval is approaching the MTTF value, the approximation is not good, and we might use the following improved approximation:

$$\lambda_A(\tau) = \left(\frac{\Gamma(1+1/\alpha)}{\text{MTTF}} \right)^{\alpha} \tau^{\alpha-1} \left[1 - 0.1\alpha(\tau/\text{MTTF})^2 + (0.09\alpha - 0.2)\tau/\text{MTTF} \right] \tag{59}$$

In MS Excel the gamma function could be found by EXP(GAMMALN(x)). (note that in national versions of Excel, the exponential function is often written in the national language, e.g. EKSP in Norwegian). If the gamma function is not available at hand Table 2 could be used to look-up the expression required in equation (58).

Table 2 $\Gamma(1+1/\alpha)$ for selected values of α

α	$\Gamma(1+1/\alpha)$	$\Gamma(1+1/\alpha)^{\alpha}$
1.5	0.903	0.858
2	0.886	0.785
2.5	0.887	0.742
3	0.893	0.712
3.5	0.900	0.691
4	0.906	0.675
4.5	0.913	0.663
5	0.918	0.653

If ageing is specified qualitatively, we may use Table 3.

Table 3 Effective failure rate as a function of maintenance interval

Situation	Effective failure rate $\lambda_A(\tau)$	Reduction
Weak ageing	$0.79\ \tau/\ \text{MTTF}^2$	$(1-0.79\ \tau/\ \text{MTTF}) \times 100\%$
Medium ageing	$0.71\ \tau^2/\ \text{MTTF}^3$	$(1-0.71\ \tau^2/\ \text{MTTF}^2) \times 100\%$
Strong ageing	$0.67\ \tau^3/\ \text{MTTF}^4$	$(1-0.67\ \tau^3/\ \text{MTTF}^3) \times 100\%$

Example 7.2

Assume we have a situation with "medium ageing", and maintenance interval $\tau = 0.3\times$MTTF. This then gives a reduction of $1\text{-}0.71\times0.3\times0.3 \approx 94\%$ compared to a situation without maintenance. □

Note that in order to utilise equation (58) we need to know the MTTF for the component if it is not maintained. In some situations we have observed data for a given maintenance strategy, say maintenance interval τ_0. Now assume that we based on the data have estimated the failure rate (number of failures divided by the service time) to be λ_0. We might then use the following approximation for the effective failure rate:

$$\lambda_A(\tau) = \lambda_0 \left(\frac{\tau}{\tau_0} \right)^{\alpha-1} \tag{60}$$

Equation (58) is in fact an approximation in the so-called block replacement policy (PRP). In the following we shortly describe both this policy as well as the age replacement policy (ARP).

Age replacement policy

In the age replacement policy (ARP) the component is replaced periodically when it reaches a fixed age. If the component fails within a maintenance interval, the component is replaced, and the "maintenance clock" is reset. Figure 34 shows a realization of an ARP, where we usually replace the component after a service time of τ. In some situations the component fails in the maintenance interval, indicated by the failure times T_1 and T_2.

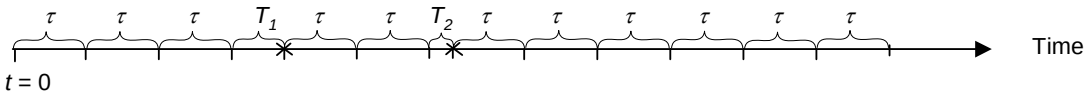


Figure 34 Realisation of an ARP

In the ARP we assume that after a replacement, the component is as good as new, and time to failure have the same distribution, say $F_T(t)$ in all maintenance intervals. We also observe that the length of a maintenance cycle (T_{MC}) is a random quantity. The expected length of a maintenance cycle is given by:

$$E(T_{MC}) = \int_0^\tau t f_T(t)\mathrm{d}t + \tau P(T > \tau) = \int_0^\tau \big(1 - F_T(t)\big)\mathrm{d}t \tag{61}$$

Further, the probability of failure within a maintenance cycle is given by $P(T<\tau) = F_T(\tau)$. The average number of failures per unit time is thus given by:

$$\lambda_{AR}(\tau) = \frac{F_T(\tau)}{\int_0^\tau \big(1 - F_T(t)\big)\mathrm{d}t} \tag{62}$$

Numerical methods are usually required to utilise Equation (62).

Block replacement policy

The block replacement policy (PRP) is similar to the ARP, but we do not reset the maintenance clock if a failure occurs in a maintenance period.

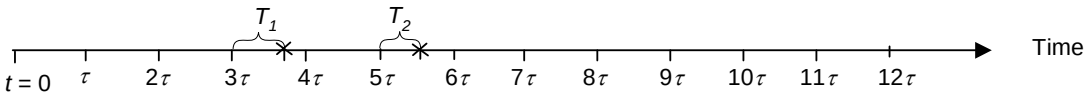


Figure 35 Realisation of a BRP

The BRP seems to be "wasting" some valuable component life time, since the component is replaced at an age lower than τ if a failure occurs in a maintenance period. However, this could be defended due to administrative savings, or reduction of "set-up" cost if many components are maintained simultaneously. Note that we have assumed that the component was replaced upon failure within one maintenance interval. In some situations a "minimal repair", or an "imperfect repair" is carried out for such failures. This will, however, not be discussed in the following where we assume a perfect repair, or replacement bringing the component to a "good as new" state. In this situation, the maintenance cycle length is fixed (τ), and the average number of failures per unit time is given by the expected number of failures in (0,t):

$$\lambda_{BR}(\tau) = \frac{\text{Expected number of failures in } [0,\tau)}{\tau} = \frac{W(\tau)}{\tau} \tag{63}$$

Where $W(t)$ is the renewal function discussed later on. If τ is small compared to the MTTF of the component, it is very unlikely to have more than one failure within one maintenance cycle. In this case $W(t) \approx F_T(t)$, and Equation (58) follows if we assume Weibull distributed times to failure.

7.3.4 Shock

In the shock model, the failures are assumed to be caused by internal or external shocks. The time from a shock occurs, until the component fails is assumed to be so short that there exist no effective maintenance tasks that could "catch" this failure. However, for systems that have a hidden function, it would be beneficial to reveal if a failure has occurred, and then repair the failure before a real demand occurs. For a component with a hidden function which is periodically tested, the probability of failure on demand (PFD) is approximately $\lambda\tau/2$. If the rate of demand is f_D, then the "effective failure rate", or more precisely the rate of undetected demands is given by:

$$\lambda_{FT}(\tau) = \frac{f_D \lambda \tau}{2} \tag{64}$$

8. STOCHASTIC POINT PROCESS

8.1 Introduction

In the previous section life times were discussed. In such a situation a component or a system is observed until the first failure is observed. What happens after a failure has not been discussed. Very often the component or system is repaired after a failure and put into service again. For such a situation the life time approach will not be appropriate. Instead it is more appropriate to work with what is called a stochastic point process. A textbook on the discussion of life time analysis versus point processes is Ascher&Feingold (1984).

In this context a *stochastic point process* is a mathematical model for highly localized events (failures) distributed randomly on the time axis. By "highly localized" is here meant that failures occur instantaneously in time. This will, however, be an approximation to the real life where a failure is thought of as a deteriorating process. Generally, the "points" could represent anything, but in our context we think about these points as failures.

Three important point processes are discussed in this section:

- The Homogeneous Poisson Process (HPP)
- The Renewal Process (RP)
- The Non Homogeneous Poisson Process (NHPP)

In the HPP we assume that failures are exponentially distributed, and after a failure we start over with a new component with the same life time distribution. In the renewal process the situation is more general, we allow the failure times to come from any life time distribution, but we still assume that after a system failure we proceed with a new component with the same life time distribution as we originally had. In the NHPP the time to the first failure could come from any life time distribution, but now we assume that after a failure the process continues with the same probability of failure as was just prior to the failure, i.e. a minimal repair situation.

8.2 Basic definition needed for stochastic point processes

Cumulative number of failures, $N(t)$: Consider a stochastic point process starting at time 0. We let $N(t)$ denote the cumulative number of failures in the time period from 0 to t.

Expected cumulative number of failures, $W(t)$: The expected cumulative number of failures in the time period from 0 to t is:

$$W(t) = E[N(t)]$$

Renewal function, $W(t)$: In a renewal process where the system is renewed after a failure the renewal function $W(t)$ is the expected number of renewals, which equals the expected number of failures, i.e.

$$W(t) = E[N(t)]$$

Rate of occurrence of failures (ROCOF): The ROCOF is the time derivative of the expected cumulative number of failures:

$$\text{ROCOF} = w(t) = \frac{dW(t)}{dt}$$

To best interpret the ROCOF, write:

$w(t)\Delta t \approx E[N(t+\Delta t) - N(t)]$ = expected number of failures in $(t, t+\Delta t)$

or in terms of probabilities:

$w(t)\Delta t \approx \Pr(\text{Failure in } (t, t+\Delta t))$

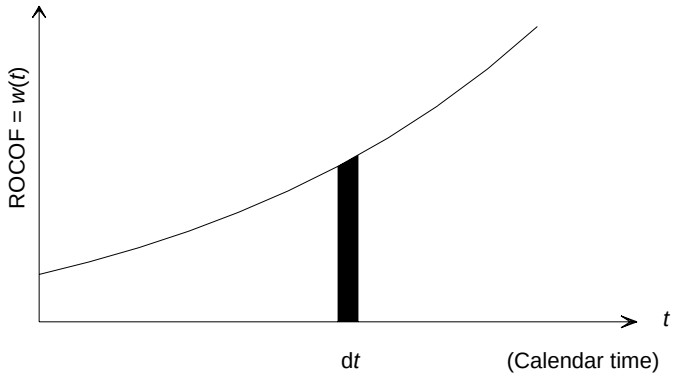


Figure 36 Interpretation of the ROCOF

Renewal rate, $w(t)$: In a renewal process where the system is renewed after a failure the renewal rate $w(t)$ is the probability of a renewal in a small time interval, which will coincide with the ROCOF:

$$w(t) = \frac{dW(t)}{dt}$$

As good as new. A component is said to be as good as new after a repair if:

a) The time to failure of a component repaired at time r, say X_r, is distributed identically to the time to failure of the original component installed at time 0, say X. i.e. $\Pr(X_r > s+r) = \Pr(X > s)$

b) X_r is independent of the history up to time r.

Minimal repair. With "minimal repair" is meant that a *system* is repaired to the state the system was just before the failure occurred. The concept of minimal repair can be a very good approximation for systems which have a large number of *components*. When one component fails this is assumed to cause a system failure. Further it is assumed that only the failed component is repaired. It may then be assumed that the total state of the system can be almost the same as just before the actual component failed. In real life, it is likely to assume something between "minimal repair" and "as good as new after a repair".

Note
There is a difference between the ROCOF and the hazard rate which is very often overlooked. The ROCOF, $w(t)$, is proportional to the *unconditional* probability of experiencing a failure in a small interval, whereas the hazard rate, $z_X(x)$ is proportional to the *conditional* probability of experiencing a failure in a small interval *given* survival up to time x. For many systems the concept of ROCOF and FOM could be combined as illustrated in Figure 37.

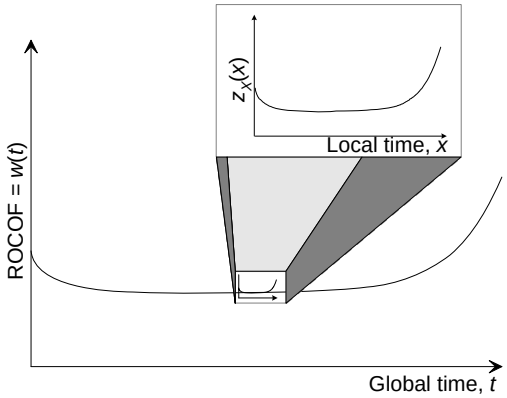


Figure 37 Global vs local time

8.3 The Homogeneous Poisson Process (HPP)

For the Homogeneous Poisson Process the situation is as follows:

- A system is put into service at time $t = 0$.
- Let X_1 denote the first system failure, X_2 denote the time from the first failure is repaired until the second failure occurs and so on.
- Repair times are assumed to be so small that they might be ignored
- The life times $X_1,X_2,\ldots$ are assumed to be independent and identically exponential distributed with parameter λ.

For such a process the following results applies:

a) The rate of occurrence of failures, ROCOF is constant and independent of time, i.e. $w(t) = \lambda$.
b) The number of failures in an interval (a,b) is Poisson distributed with parameter $\lambda(b-a)$, i.e. $\Pr(N(b) - N(a) = n) = \dfrac{\lambda(b-a)}{n!} e^{-\lambda(b-a)}$
c) The mean number of failures in an interval (a,b) is $E(N(b) - N(a)) = \lambda\,(b-a)$.
d) The renewal function is $W(t) = \lambda t$

8.4 The renewal process (RP)

The following situation applies for the renewal process

- A system is put into service at time $t = 0$.
- Let X_1 denote the first system failure, X_2 denote the time from the first failure is repaired until the second failure occurs and so on.

- Repair times are assumed to be so small that the might be ignored
- The life times $X_1, X_2, \ldots$ are assumed to be independent and identically distributed. However, we do not longer restrict to the exponential distribution. Very often the Weibull distribution is considered.

The renewal process if of great importance for the "Block Replacement Policy" used in maintenance. Consider the following situation:

- A component is installed at time $t = 0$.
- Regardless of the number of failures, and when they occurred, the component is replaced with a new one at time τ.
- If the component fails before time τ, the component is replaced with a new component with the same life time distribution as the original component.
- If there is a second component failure before τ, this one is also replaced with a new one and so forth.
- Component failures are assumed to be independent, and repair time could be neglected.
- The sequence is repeated after τ.

It is now easy to verify that this situation matches with a renewal process. Of interest will be the renewal function, $W(\tau)$, i.e. the expected number of (unplanned) failures in the interval (0, τ). Usually the maintenance interval τ is much less than the mean time to failure (MTTF) of the components.

The renewal function could be expressed by (see e.g. Rausand and Høyland 2004):

$$W(t) = \sum_{r=1}^{\infty} F_X^{(r)}(t) \tag{65}$$

where $F_X^{(r)}(t)$ is the r'th convolution of the distribution of X. Remember that $F_X(x)$ is the distribution function of the failure times. Further if $X_1, X_2, \ldots$ are such failure times, then the r'th convolution could be found by:

$$F_X^{(r)}(t) = \Pr(\sum_{i=1}^{r} X_i \le t) \tag{66}$$

In maintenance applications we usually want to calculate $W(t)$ for small values of t. This means that $F_X^{(r)}(t)$ would very fast approach to zero as r increases. The summation in Equation (65) will therefore converge very fast. The challenge is thus to calculate the convolution given by Equation (66) for a small number of values of r. Generally we could find the r'th convolution by the following recursive formula:

$$F_X^{(r)}(t) = \int_0^t F_X^{(r-1)}(t-u) f_X(u) du \tag{67}$$

where $F_X^{(1)}(t) = F_X(t)$. Thus, if we have access to a routine for numerical integration, we could easily calculate the r'th convolution, and sum them up to get the renewal function $W(t)$. A slightly more sophisticated approach would be to use the renewal equation:

$$W(t) = F_X(t) + \int_0^t W(t-u) f_X(u) du \tag{68}$$

Now, if we have an initial estimate $W_0(t)$ of the renewal function, the following iterative formula could be used to improve the estimate of the renewal function:

$$W_i(t) = F_X(t) + \int_0^t W_{i-1}(t-u) f_X(u) du \tag{69}$$

For small values of t an intuitive initial estimate is $W_0(t)=F_X(t)$.

If we assume Weibull distributed failure times the essential source code in Visual Basic is:

```
dt = t / nMax
For i = 0 To nMax
    W(i) = 1# - Exp(-(lambda * i * dt) ^ alpha)
    F1(i) = W(i)
    f(i) = alpha * lambda * (lambda * i * dt) ^ (alpha - 1) * (1 - W(i))
Next
Do
    PrevW = W(nMax)
    For i = nMax To 1 Step -1
        W(i) = F1(i) + IntConv(i, dt, W, f)
    Next
Loop While Abs(PrevW - W(nMax)) / PrevW > eps
Return W(nMax)
```

Where `W(0:nMax)`, `F1(0:nMax)` and `f(0:nMax)` are arrays containing $W(t)$, $F_X(t)$ and $f_X(t)$. `nMax` is the number of steps used in the numerical integration, and `IntConv()` is a function that performs numerical integration of the convolution of $W(u)$ and $f_X(u)$ from 0 up to $i \times dt$. `eps` is the required precision, e.g. 1e-3. The iteration scheme will converge after two or three iterations for reasonable large values of the ageing parameter ($\alpha > 1.5$) and small value of τ (τ < MTTF).

To calculate the effective failure rate of an ageing component, $\lambda_A(\tau)$, we may thus set:

$$\lambda_A(\tau) = W(\tau)/\tau \tag{70}$$

A MS Excel program (WeibullRenwal.xls) is available for the calculation of $W(\tau)$ and $\lambda_A(\tau) = W(\tau)/\tau$ in the Weibull situation.

8.5 The Non Homogeneous Poisson Process (NHPP)

The following situation applies for non homogeneous Poisson process

- A system is put into service at time $t = 0$.
- If the system fails, a repair is conducted and the system is put into service after a time that could be neglected
- The repair action set the system back to a state as good as it was immediately prior to the failure, i.e. a minimal repair.

For such a process the following results apply:

a) The rate of occurrence of failures, ROCOF = $w(t)$ is generally not constant.
b) The number of failures in an interval (a,b) is Poisson distributed with parameter

$\lambda = \int_a^b w(t)dt$, i.e. $P(N(b) - N(a) = n) = \frac{\int_a^b w(t)dt}{n!} e^{-\int_a^b w(t)dt}$

c) The mean number of failures in an interval (a,b) is $E(N(b) - N(a)) = \int_a^b w(t)dt$

d) The cumulative number of failures up to time t is $W(t) = \int_0^t w(u)du$

We will briefly summarize the main results for three parametric NHPP models:

- The power law model
- The linear model
- The log-linear model

Table 4 Properties for selected NHPP models

↓Property Model→	Power law model	Linear model	Log-linear model
ROCOF = $w(t)$	$\lambda\beta t^{\beta-1}$	$\lambda(1+\alpha t)$	$e^{\alpha+\beta t}$
$W(t)$	λt^{β}	$\lambda(t+\alpha t^2/2)$	$(e^{\alpha+\beta t} - e^{\alpha})/\beta$
System improves for	$\beta < 1$	$\alpha < 0$	$\beta < 0$
System deteriorates for	$\beta > 1$	$\alpha > 0$	$\beta > 0$
Average failure rate when replaced at time τ	$\lambda\tau^{\beta-1}$	$\lambda(1+\alpha\tau/2)$	$(e^{\alpha+\beta\tau} - e^{\alpha})/(\beta\tau)$

9. STRUCTURE FUNCTION AND SYSTEM RELIABILITY

In this chapter we will present some simple methods for analysing system comprised of several components, where the reliability performance of each component is known, i.e. in terms of failure rates and repair times.

For components we have

$$x(t) = \begin{cases} 1 \text{ if the component is functioning at time } t \\ 0 \text{ if the component is in a fault state at time } t \end{cases} \tag{71}$$

For the system we now introduce

$$\phi(\mathbf{x}, t) = \begin{cases} 1 \text{ if the system is functioning at time } t \\ 0 \text{ if the system is in a fault state (not functioning) at time } t \end{cases} \tag{72}$$

ϕ denotes the structure function, and depends on the x_is ($\boldsymbol{x}$ is a vector of all the x_is). $\phi(\boldsymbol{x},t)$ is thus a mathematical function that uniquely determines whether the system functions or not for a given value of the $\boldsymbol{x}$-vector. Not it is not always straight forward to find a mathematical expression for $\phi(\boldsymbol{x},t)$.

9.1 Reliability Block Diagram (RDB)

Reliability block diagrams are valuable when we want to visualise the performance of a system comprised of several (binary) components.

Figure 38 and Figure 39 shows the reliability block diagram for simple structures. The interpretation of the diagram is that the system is functioning if it is a connection between *a* and *b*, i.e. it is a path of functioning components from *a* to *b*. The system is in a fault state (is not functioning) if it does not exist a path of functioning components between *a* and *b.*.

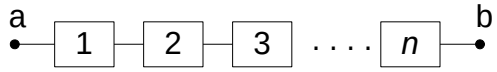


Figure 38 Reliability block diagram for a serial structure

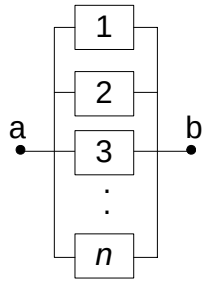


Figure 39 Reliability block diagram for a parallel structure

9.2 The structure function for some simple structures

For a serial function we have:

$$\phi(\mathbf{x}) = x_1 \cdot x_2 \cdot \ldots \cdot x_n \tag{73}$$

For a parallel structure we have

$$\phi(\mathbf{x}) = 1-(1-x_1)(1-x_2)\ldots(1-x_n) \tag{74}$$

For a *k*-out-of-*n* structure we have

$$\phi(\mathbf{x}) = \begin{cases} 1 \text{ if } \sum_{i=1}^{n} x_i \geq k \\ 0 \text{ if } \sum_{i=1}^{n} x_i < k \end{cases} \tag{75}$$

A *k*-out-of-*n* system is a system that functions if and only if at least *k* out of the *n* components in the system is functioning. We often write *k* oo *n to* denote a *k* out of *n* system, for example 2 oo 3,

For structures comprised of serial and parallel structures we can combined the above formulas as.

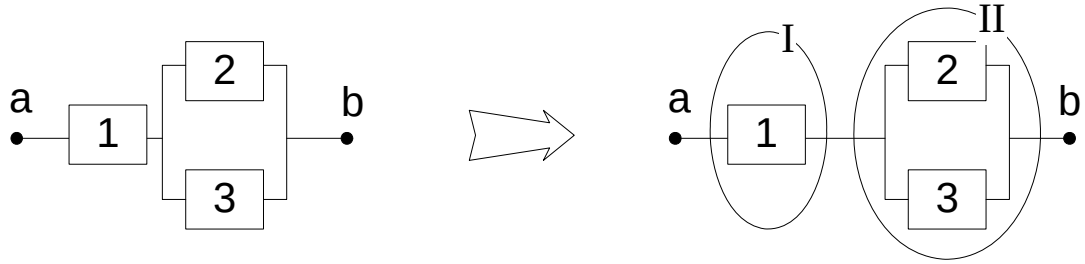


Figure 40 Splitting the reliability block diagram in sub-blocks

Figure 40 shows how we may split the reliability block diagram into sub-blocks, here I and II. We may then write $\phi(x) = \phi_I \times \phi_{II}$ because I and II is in serial. Further, we have $\phi_I = x_1$, and $\phi_{II} = 1-(1-x_2)(1-x_3)$, thus we have $\phi(x) = x_1(1-(1-x_2)(1-x_3))$.

Exercise 11
Verify equation (75) for a 2 oo 3 system. □

9.3 Using the structure function

9.3.1 System reliability

If we have a mathematical expression for the structure function, and we know the values of the x_is (component states), we could determine if the system is functioning by "inserting" the *x*-values into the structure function.

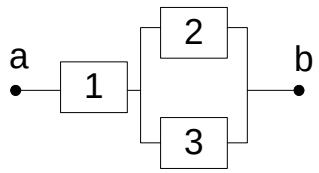


Figure 41 Simple reliability block diagram

For the system in Figure 41 we easily see that the structure function is given by $\phi(\boldsymbol{x}) = x_1(1-(1-x_2)(1-x_3))$, and by inserting for example $x_1 = 1$, $x_2 = x_3 = 0$, we find $\phi(\boldsymbol{x}) = 1(1-(1-0)(1-0)) = 0$, that is the system is not functioning.

Further, if we know the structure function, and we know the component reliabilities (p=1-U) we may insert the p_is for the x_is in order to obtain the system reliability (notation: $h(\mathbf{p})$). From the previous example we have:

$$h(\mathbf{p}) = p_1(1-(1-p_2)(1-p_3)) \tag{76}$$

where p_1, p_2, and p_3 are the component reliabilities. If we for example let $p_1 = 0.9$, $p_2 = 0.8$, and $p_3 = 0.7$ tis gives a system reliability of $h(\mathbf{p}) = 0.9\times(1-0.2\times0.3) = 0.9\times0.94 = 0.846$.

The method we presented is only valid if:

- The components are stochastically independent
- We have "multiplied out" the structure function, and removed any powers, e.g. x_i^n is replaced with x_i.

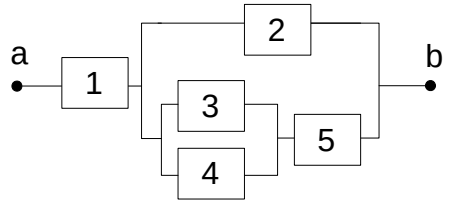
Thus, the following procedure may be used for calculating system reliability $h(\mathbf{p})$ when the components are independent.

1. Obtain the structure function $\phi(\boldsymbol{x})$
2. Multiply out all terms in $\phi(\boldsymbol{x})$
3. Remove all exponents in powers of x, i.e. replace x_i^n with x_i for $n > 1$. Denote the result $\phi^M(\boldsymbol{x})$
4. The system reliability is found by replacing the x_i's in $\phi^M(\boldsymbol{x})$ with the corresponding p_i's, i.e.

$$h(\mathbf{p}) = \phi^M(\boldsymbol{x} \mid \boldsymbol{x} = \mathbf{p}) \tag{77}$$

It could be very cumbersome to "multiply out" the structure function, and in fact a computer will also fail to do this for large systems. This call for approximation formulas. A first approximation will be to use $h(\mathbf{p}) = \phi(\boldsymbol{x} \mid \boldsymbol{x} = \mathbf{p})$. A slightly better approach for hand calculation would be to "investigate" the reliability block diagram. We do that by searching for multiple occurrence of components. If a component occurs several times, and we are able to separate all occurrence of a component within one "sub-block" (see Figure 40) we could "multiply out" this sub-block and remove exponents in that sub-block. We now let $\phi^{M1}(\boldsymbol{x})$ denote the structure function where we have "resolved" one such block. We could then proceed with the next block and find $\phi^{M2}(\boldsymbol{x})$ and so on. There is, however no guarantee that we could isolate all occurrence of a component in one single sub-block. The final approximation for the system reliability will then be $h(\mathbf{p}) = \phi^{Mn}(\boldsymbol{x} \mid \boldsymbol{x} = \mathbf{p})$ if we resolve n sub-blocks.

Exercise 12
Find the structure function for the following system:



Assume that the system reliabilities are given by: $p_1 = 0.99$, $p_2 = p_5 = 0.95$, and $p_3 = p_4 = 0.9$. Find the system reliability. □

9.3.2 A measure for criticality importance

There exist several measures for criticality importance of the components in a system. We will here present Birnbaums measure for reliability importance of a component (denoted *i*):

$$I^B(i) = \partial h(\mathbf{p}) / \partial p_i \tag{78}$$

If we sort the components according to their reliability importance, this is a good starting point for:

- Eventually replacing components with higher reliability
- Prioritisation of maintenance resources.

For simple structure functions we could obtain Brinbaums measure analytically. However, for more complex systems we need to obtain Birnbaums measure numerically. We may then use the following result:

$$I^B(i) = h(\mathbf{p} \mid p_i = 1) - h(\mathbf{p} \mid p_i = 0) \tag{79}$$

Where $h(\mathbf{p} \mid p_i = u)$ is the value of $h(\mathbf{p})$ when $p_i = u$.

Exercise 13

Find Birnbaums measure of reliability importance for a serial structure, and a parallel structure. □

Exercise 14

Consider the system in Exercise **12**, and find Birnbaums measure of reliability importance for all components. □

9.3.3 Frequency of system failures, F_0

In Section 9.3.1 we presented a procedure to calculate the system reliability of a reliability block diagram. $h(\mathbf{p})$ is the probability that the system is functioning, and similarly $U = 1 - h(\mathbf{p})$ is the probability that the system is not functioning, i.e. the unavailability of the system. In some situations we would also like to calculate the *frequency* of system failures, F_0. By using the fact that Birnbaums measure of reliability importance also could be interpreted as the probability that component *i* is critical, we realise that:

$$F_0 = \sum_i I^B(i) \times \lambda_i \tag{80}$$

where λ_i is the failure rate of component *i.*

Example 9.1 - Optimisation of maintenance of a component in an RBD

We will consider the reliability block diagram in Exercise **12**. To carry out the calculations, we use the Excel program RBDUtil.xls. We first enter the reliability data as given in Exercise

12 into the $p(i)$ column of the spreadsheet. Further we enter the number of components (No. Comp) to 5. In the $h(p)$ cell we enter[6] the structure function as:

```
=p_01 * ip( p_02, ip( p_03, p_04 ) * p_05 )
```

Note that RBDUtil.xls supports the "ip" function, where

$$\text{ip}(p_1, p_2, \ldots, p_n) = 1 - (1- p_1)(1- p_2)\ldots(1- p_n)$$

With the reliability data we entered, we obtain $h(\mathbf{p})$=0.9871. RBDUtil.xls also calculates Birnbaums measure of reliability, and the result is shown in the $I^B(i)$-column.

We will no consider the maintenance of component 2. Assume that without any preventive maintenance, MTTF_N for component 2 is one month (730 hours). Further assume a mean down time (MDT) of 39 hours which gives $p_2 = \text{MTTF}_N/(\text{MTTF}_N+\text{MDT}) = 0.95$. If we have ageing failures, we could reduce the effective failure rate by replacing the component with a new one at predetermined intervals of length τ. We will assume the ageing parameter α to be 2.5. There are four cost elements to be included in the cost model:

$\text{PM}_{\text{Cost}}(2) =$ 1 000 = cost per preventive maintenance action (component 2)
$\text{CM}_{\text{Cost}}(2) =$ 5 000 = cost per corrective maintenance action (component 2)
SF_{Cost} = 20 000 = cost per system failure
U_{Cost} = 10 000 = cost per time unit (hour) when the system is unavailable

The "contribution" to the total cost for maintenance and operation of the system with respect to component 2 is (cost per unit time):

$$C(\tau) = C_{\text{PM}}(\tau) + C_{\text{CM}}(\tau) + C_{\text{SF}}(\tau) + C_U(\tau)$$

$$= \text{PM}_{\text{Cost}}(2)/\tau + \lambda_E(\tau)[\text{CM}_{\text{Cost}}(2) + I^B(2) \times \text{SF}_{\text{Cost}}] + (1-p_2) \times I^B(2) \times U_{\text{Cost}} \qquad (81)$$

The RBDUtil.xls summarises the total cost contribution according to Equation (81) for each component, and the result is shown for component 2 in Figure 42. The cost is minimised for τ approximate equal to 350 hours, i.e. the preventive maintenance activity should be carried out every 14th day.

No. Comp	5					Cost per system failure			20 000						
$h(p)$	0.9893					Downtime Cost per time unit			10 000						
Comp #	$p(i)$	$I^B(i)$	MTTF	α	MDT	τ	$\lambda_A(\tau)$	PM_{Cost}	CM_{Cost}	$C_{PM}(\tau)$	$C_{CM}(\tau)$	$C_{Trip}(\tau)$	$C_U(\tau)$	$C_{Tot}(\tau)$	
1	0.99	0.98705477								0	0	0	98.705	98.70548	
2	0.9875	0.05595976	730	2.5	39	350	0.0003	1000	5000	2.8571	1.60817	0.35997	5.596	10.42126	
3	0.9	0.00423229								0	0	0	0.4232	0.423229	
4	0.9	0.00423229								0	0	0	0.4232	0.423229	
5	0.95	0.0465548								0	0	0	4.6555	4.65548	

Figure 42 Calculation result with RBDUtil.xls

Note that we have supported the WeiFreq() function in RBDUtil.xls, where WeiFreq() corresponds to $\lambda_E(\tau) = W(\tau)/\tau$ in the situation where we have Weibull distributed failure times. □

[6] In MS Excel we could enter a formula into a cell by prefixing it by an equal sign.

10. RELIABILITY CENTRED MAINTENANCE

Reliability centred maintenance (RCM) is a method for maintenance planning developed within the aircraft industry and later adapted to several other industries and military branches. A major advantage of the RCM methodology is a structured, and traceable approach to determine type of preventive maintenance. This is achieved through an explicit consideration of failure modes and failure causes. A major challenge in an RCM analysis is to limit the scope of the analysis so that it is possible to carry out the analysis within the limits of time and budget. Most implementations of RCM put main focus on the identification of maintenance tasks, but do not carry out explicit optimisation of maintenance intervals. We will, however, present an approach to RCM that also enables optimisation of maintenance intervals. In order to do so, we need to structure the analysis much more than what is common in most RCM approaches.

Structuring take place at several steps in the RCM analysis. Because the failure mode and effect analysis (FMECA) is very time consuming, and because the basis for maintenance optimisation also is established through the FMECA we will introduce several means to simplify and structure this part of the analysis:

- Introduction of so-called TOP-events in the analysis. Such a TOP event could be "derailment", "fire", "collision train-train" for safety, and "Slow speed –40 km/h" and "Full stop" etc for punctuality. For these identified TOP events a general assessment is carried out where the total risk or cost for each such TOP event is "calculated". The "consequence" analysis is thus reduced to totally 10-15 items, which is a very low number compared to the number of "rows" in the FMECA, which could be thousands or more.
- Introduction of generic RCM templates. A generic RCM template is the result of a general analysis of an equipment such as a turnout (mechanical part), a switch motor (electrical part), the traction system of a train etc. In such a generic analysis we make an "average" assessment of important reliability parameters. Experience has shown that the number of "generic" RCM templates is in the order of 50, where each generic template comprise 5 to 10 "components".
- When the maintenance program is established for a specific line, or a specific train set, the generic RCM template is taken as a starting point. For this general template we make local adjustment in terms of adjustment factors. When the local adjustment factors have been defined, it is straight forward to "update" the generic template to a local analysis, where the optimisation of maintenance intervals also could be automated.
- When we know that we have several hundred thousand physical components to treat when the maintenance program is defined, we can imagine the value of such a "generic" and "local adjustment" approach.

The RCM analysis may be carried out as a sequence of activities. Some of these activities, or steps, are overlapping in time. The RCM process comprises the following steps:

1. Study preparation
2. System selection and definition
3. Functional failure analysis (FFA)
4. Critical item selection
5. Data collection and analysis
6. Failure modes, effects and criticality analysis (FMECA)
7. Selection of maintenance actions
8. Determination of maintenance intervals

9. Preventive maintenance comparison analysis
10. Treatment of non–critical items
11. Implementation
12. In–service data collection and updating
13. Local adjustments

The various steps are discussed in the following sections with a focus on Steps 1–8. Note that the basis for step 1-12 would be the "generic approach". That is, we typically carry out these steps for "generic" systems or components, and then in step 13 we make explicit assessments reflecting the conditions related to each physical unit.

10.1 Step 1: Study preparation

The main objectives of an RCM analysis are:

1. to identify effective maintenance tasks,
2. to evaluate these tasks by some cost–benefit analysis, and
3. to prepare a plan for carrying out the identified maintenance tasks at optimal intervals.

If a maintenance program already exists, the result of an RCM analysis will often be to eliminate inefficient maintenance tasks.

Before an actual RCM analysis is initiated, an RCM project group should be established, see e.g. Moubray (1991) pp. 16–17. The RCM project group should include at least one person from the maintenance function and one from the operations function, in addition to an RCM specialist.

In Step 1 "Study preparation" the RCM project group should define and clarify the objectives and the scope of the analysis. Requirements, policies, and acceptance criteria with respect to safety and environmental protection should be made visible as boundary conditions for the RCM analysis.

The part of the plant to be analysed is selected in Step 2. The type of consequences to be considered should, however, be discussed and settled on a general basis in Step 1. Possible consequences to be evaluated may comprise:

(i) risk to humans,
(ii) environmental damages,
(iii) delays and cancellation of travels,
(iv) material losses or equipment damage,
(v) loss of marked shares, etc.

The possible consequence classes can not be measured in one common unit. It is therefore necessary to prioritise between means affecting the various consequence classes. Such a prioritisation is not an easy task and will not be discussed in this presentation. The trade–off problems can to some extent be solved within a decision theoretical framework (Vatn 1995 and Vatn *et al*. 1996).

RCM analyses have traditionally concentrated on PM strategies. It is, however, possible to extend the scope of the analysis to cover topics like corrective maintenance strategies, spare part inventories, logistic support problems, etc. The RCM project group must decide what should be part of the scope and what should be outside.

The resources that are available for the analysis are usually limited. The RCM group should therefore be sober with respect to what to look into, realizing that analysis cost should not dominate potential benefits.

In many RCM applications the plant already has effective maintenance programs. The RCM project will therefore be an upgrade project to identify and select the most effective PM tasks, to recommend new tasks or revisions, and to eliminate ineffective tasks. Then apply those changes within the existing programs in a way that will allow the most efficient allocation of resources.

When applying RCM to an existing PM program, it is best to utilise, to the greatest extent possible, established plant administrative and control procedures in order to maintain the structure and format of the current program. This approach provides at least three additional benefits:

(i) It preserves the effectiveness and successfulness of the current program.
(ii) It facilitates acceptance and implementation of the project's recommendations when they are processed.
(iii) It allows incorporation of improvements as soon as they are discovered, without the necessity of waiting for major changes to the PM program or analysis of every system.

Since we are heading for a sound basis for interval optimisation, we will need an explicit quantification of the risk associated with each "TOP event". On a general basis, we therefore need to establish the relevant risk models, both with respect to safety and punctuality. See Chapter 11 for a preliminary assessment of these risks. It is not the maintenance department that is responsible for establishing these "generic" risk models. Usually risk analyses, or safety cases exists, and these could be used as a basis for the appropriate structuring of the risk picture.

10.2 Step 2: System selection and definition

Before a decision to perform an RCM analysis is taken, two questions should be considered:

- To which systems are an RCM analysis beneficial compared with more traditional maintenance planning?
- At what level of assembly (plant, system, subsystem …) should the analysis be conducted?

Regarding the first question, all systems may in principle benefit from an RCM analysis. With limited resources, we must, however, usually make priorities, at least when introducing the RCM approach for the first time. We should start with the systems that we assume will benefit most from the analysis. The following criteria may be used to prioritise systems for an RCM analysis:

(i) The failure effects of potential system failures must be significant in terms of safety, environmental consequences, production loss, or maintenance costs.
(ii) The system complexity must be above average.
(iii) Reliability data or operating experience from the actual system, or similar systems, should be available.

Most operating plants have developed an assembly hierarchy, i.e. an organization of the system hardware elements into a structure that looks like the root system of a tree. In the offshore oil and gas industry this hierarchy is usually referred to as the tag number system. Several other names are also used. Moubray (1991) for example refers to the assembly

hierarchy as the plant register. In railway infrastructure maintenance it is common to use the disciplinary areas as the highest level in the plant register, typically we have:

- Superstructure
- Substructure
- Signalling
- Telecommunications
- Power supply (overhead line with supporting systems)
- Low voltage systems

For the rolling stock we similarly have a system breakdown:

- The breaking system including automatic train protection (ATP)
- The traction system
- The door system with interlocking connections to traction system
- The pantograph with supporting system
- The bogie system
- The coupler system
- The wagon
- The locomotive

The following terms will be used in this paper for the levels of the assembly hierarchy:

Plant: A logical grouping of systems that function together to provide an output or product by processing and manipulating various input raw materials and feed stock. An offshore gas production platform may e.g. be considered as a plant. For railway application a plant might be a maintenance area, where the main function of that "plant" is to ensure satisfactiory infrastructure functionality in that area. Moubray (1991) refers to the plant as a cost center. In railway application a plant corresponds to a train set (rolling stock), or a line (infrastructure).

System: A logical grouping of subsystems that will perform a series of key functions, which often can be summarized as one main function, that are required of a plant (e.g. feed water, steam supply, and water injection). The compression system on an offshore gas production platform may e.g. be considered as a system. Note that the compression system may consist of several compressors with a high degree of redundancy. Redundant units performing the same main function should be included in the same system. It is usually easy to identify the systems in a plant, since they are used as logical building blocks in the design process.

The system level is usually recommended as the starting point for the RCM process. This is further discussed and justified for example by Smith (1993) and in MIL–STD 2173. This means that on an offshore oil/gas platform the starting point of the analysis should be for example the compression system, the water injection system or the fire water system, and not the whole platform. In railway application the systems were defined above as the highest level in the plant hierarchy.

The systems may be further broken down in subsystems, and subsubsystems, etc. For the purpose of the RCM–process the lowest level of the hierarchy should be what we will call an RCM analysis item:

RCM analysis item: A grouping or collection of components which together form some identifiable package that will perform at least one significant function as a stand–alone item (e.g. pumps, valves, and electric motors). For brevity, an RCM analysis item will in the

following be called an analysis item. By this definition a shutdown valve, for example, is classified as an analysis item, while the valve actuator is not. The actuator is a supporting equipment to the shutdown valve, and only has a function as a part of the valve. The importance of distinguishing the analysis items from their supporting equipment is clearly seen in the FMECA in Step 6. If an analysis item is found to have no significant failure modes, then none of the failure modes or causes of the supporting equipment are important, and therefore do not need to be addressed. Similarly if an analysis item has only one significant failure mode then the supporting equipment only needs to be analyzed to determine if there are failure causes that can affect that particular failure mode (Paglia *et al*. 1991). Therefore only the failure modes and effects of the analysis items need to be analysed in the FMECA in Step 6. An analysis item is usually repairable, meaning that it can be repaired without replacing the whole item. In the offshore reliability database OREDA (2002) the analysis item is called an equipment unit. The various analysis items of a system may be at different levels of assembly. On an offshore platform, for example, a huge pump may be defined as an analysis item in the same way as a small gas detector. If we have redundant items, e.g. two parallel pumps, each of them should be classified as analysis items.

When we in Step 6 of the RCM process identify causes of analysis item failures, we will often find it suitable to attribute these failure causes to failures of items on an even lower level of indenture. The lowest level is normally referred to as components.

Component: The lowest level at which equipment can be disassembled without damage or destruction to the items involved. Smith (1993) refers to this lowest level as Least Replaceable Assembly (LRA), while OREDA (1997) uses the term maintainable item.

It is very important that the analysis items are selected and defined in a clear and unambiguous way in this initial phase of the RCM–process, since the following analysis will be based on these analysis items. If the OREDA database is to be used in later phases of the RCM process, it is recommended as far as possible to define the analysis items in compliance with the "equipment units" in OREDA.

10.3 Step 3: Functional failure analysis (FFA)

The objectives of this step are:

i) to identify and describe the systems' required functions,
ii) to describe input interfaces required for the system to operate, and
iii) to identify the ways in which the system might fail to function.

Step 3(i): Identification of system functions

The objective of this step is to identify and describe all the required functions of the system.

In many guidelines and textbooks (e.g. Cross 1994), it is recommended that the various functions are expressed in the same way, as a statement comprising a verb plus a noun – for example, "close flow", "contain fluid", "transmit signal".

A complex system will usually have a high number of different functions. It is often difficult to identify all these functions without a checklist. The checklist or classification scheme of the various functions presented below may help the analyst in identifying the functions. The same scheme will be used in Step 6 to identify functions of analysis items. The term item is therefore used in the classification scheme to denote either a system or an analysis item.

1. *Essential functions:* These are the functions required to fulfil the intended purpose of the

item. The essential functions are simply the reasons for installing the item. Often an essential function is reflected in the name of the item. An essential function of a pump is for example to pump a fluid.

2. *Auxiliary functions:* These are the functions that are required to support the essential functions. The auxiliary functions are usually less obvious than the essential functions, but may in many cases be as important as the essential functions. Failure of an auxiliary function may in many cases be more critical than a failure of an essential function. An auxiliary function of a pump is for example containment of the fluid.
3. *Protective functions:* The functions intended to protect people, equipment and the environment from damage and injury. The protective functions may be classified according to what they protect, as:
   - safety functions
   - environment functions
   - hygiene functions

   Safety protective functions are further discussed e.g. by Moubray (1991) pp. 40–42. An example of a protective function is the protection provided by a rupture disk on a pressure vessel (e.g. a separator).
4. *Information functions:* These functions comprise condition monitoring, various gauges and alarms etc.
5. *Interface functions:* These functions apply to the interfaces between the item in question and other items. The interfaces may be active or passive. A passive interface is for example present when an item is a support or a base for another item.
6. *Superfluous functions:* According to Moubray (1991) "Items or components are sometimes encountered which are completely superfluous. This usually happens when equipment has been modified frequently over a period of years, or when new equipment has been over specified". Superfluous functions are sometimes present when the item has been designed for an operational context that is different from the actual operational context. In some cases failures of a superfluous function may cause failure of other functions.

For analysis purposes the various functions of an item may also be classified as:

(a) *On–line functions:* These are functions operated either continuously or so often that the user has current knowledge about their state. The termination of an on–line function is called an evident failure.

(b) *Off–line functions:* These are functions that are used intermittently or so infrequently that their availability is not known by the user without some special check or test. The protective functions are very often off–line functions. An example of an off–line function is the essential function of an emergency shutdown (ESD) system on an oil platform. Many of the protective functions are off-line functions. The termination of an off–line function is called a hidden failure.

Note that this classification of functions should only be used as a checklist to ensure that all relevant functions are revealed. Discussions about whether a function should be classified as "essential" or "auxiliary" etc. should be avoided. Also note that the classification of functions here is used at the system level. Later the same classification of functions is used in the failure modes, effects and criticality analysis (FMECA) in Step 6 at the analysis item level.

The system may in general have several operational modes (e.g. running, and standby), and several functions for each operating state.

The essential functions are often obvious and easy to establish, while the other functions may be rather difficult to reveal.

Step 3(ii): Functional block diagrams

The various system functions identified in Step 3(i) may be represented by functional diagrams of various types. The most common diagram is the so–called functional block diagram. A simple functional block diagram of a pump is shown in Figure 43.

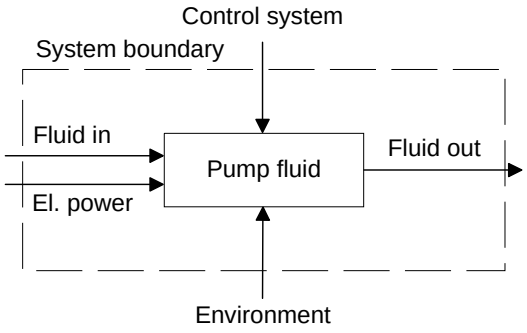


Figure 43 Functional block diagram for a pump

The necessary inputs to a function are illustrated in the functional block diagram together with the necessary control signals and the various environmental stressors that may influence the function.

It is generally not required to establish functional block diagrams for all the system functions. The diagrams are, however, often considered as efficient tools to illustrate the input interfaces to a function. The functional block diagram is recommended for RCM by Smith (1993). A detailed description of this type of diagrams is given by e.g. Pahl and Beitz (1984).

In some cases we may want to split system functions into subfunctions on an increasing level of detail, down to functions of analysis items. The functional block diagrams may be used to establish this functional hierarchy in a pictorial manner, illustrating series–parallel relationships, possible feedbacks, and functional interfaces (Blanchard & Fabrycky 1981). Alternatives to the functional block diagram are reliability block diagrams and fault trees.

Functional block diagrams are also recommended by IEC 60812 as a basis for failure modes, effects and criticality analysis (FMECA) and will therefore be a basis for Step 6 in the RCM procedure.

Step 3(iii): System failure modes

The next step of the FFA is to identify and describe how the various system functions may fail.

Since we will need the following concepts also in the FMECA in Step 6, we will use the term item to denote both the system and the analysis items. According to accepted standards (IEC 50(191)) failure is defined *as "the termination of the ability of an item to perform a required function*".

British Standard BS 5760, Part 5 defines failure mode as *"the effect by which a failure is observed on a failed item"*. It is important to realize that a failure mode is a manifestation of the failure as seen from the outside, i.e. the termination of one or more functions.

In most of the RCM references the system failure modes are denoted functional failures.

Failure modes may be classified in three main groups related to the function of the item:

i) *Total loss of function*: In this case a function is not achieved at all, or the quality of the function is far beyond what is considered as acceptable.
ii) *Partial loss of function*: This group may be very wide, and may range from the nuisance category almost to the total loss of function.
iii) *Erroneous function*: This means that the item performs an action that was not intended, often the opposite of the intended function.

A variety of classifications schemes for failure modes have been published. Some of these schemes, e.g. Blache & Shrivastava (1994), may be used in combination with the function classification scheme in Step 3(ii) to secure that all relevant system failure modes (functional failures) are identified.

The system failure modes (functional failures) may be recorded on a specially designed FFA-form, that is rather similar to a standard FMECA form. An example of an FFA-form is presented in Figure 44

System: Performed by:
Ref. drawing no.: Date: Page: of:

Operational mode	Function	Function requirements	System failure mode	Criticality			
				S	E	A	C

Figure 44 Example of an FFA-form

In the first column of Figure 44 the various operational modes of the system are recorded. For each operational mode, all the relevant functions of the system are recorded in column 2. The performance requirements to the functions, like target values and acceptable deviations are listed in column 3. For each system function (in column 2) all the relevant system failure modes are listed in column 4. In column 5 a criticality ranking of each system failure mode (functional failure) in that particular operational mode is given. The reason for including the criticality ranking is to be able to limit the extent of the further analysis by disregarding insignificant system failure modes. For complex systems such a screening is often very important in order not to waste time and money.

The criticality ranking depends on both the frequency/probability of the occurrence of the system failure mode, and the severity of the failure. The severity must be judged at the plant level.

The severity ranking should be given in the four consequence classes; (S) safety of personnel, (E) environmental impact, (A) production availability, and (C) economic losses. For each of these consequence classes the severity should be ranked as for example (H) high, (M) medium, or (L) low. How we should define the borderlines between these classes, will depend on the specific application.

If at least one of the four entries are (M) medium or (H) high, the severity of the system failure mode should be classified as significant, and the system failure mode should be subject to further analysis.

The frequency of the system failure mode may also be classified in the same three classes. (H) high may for example be defined as more than once per 5 years, and (L) low less than once per 50 years. As above the specific borderlines will depend on the application.

The frequency classes may be used to prioritise between the significant system failure modes.

If all the four severity entries of a system failure mode are (L) low, and the frequency is also (L) low, the criticality is classified as insignificant, and the system failure mode is disregarded in the further analysis. If, however, the frequency is (M) medium or (H) high the system failure mode should be included in the further analysis even if all the severity ranks are (L) low, but with a lower priority than the significant system failure modes.

In Section 15.3 we have shown a much simpler approach to the functional failure analysis than described above. Such an approach to functional failure analysis was taken in the RCM project of the Norwegian Railway Administration (Jernbaneverket).

10.4 Step 4: Critical item selection

The objective of this step is to identify the analysis items that are potentially critical with respect to the system failure modes (functional failures) identified in Step 3(iii). These analysis items are denoted functional significant items (FSI). Note that some of the less critical system failure modes have been disregarded at this stage of the analysis. Further, the two failure modes “total loss of function” and “partial loss of function” will often be affected by the same items (FSIs).

For simple systems the FSIs may be identified without any formal analysis. In many cases it is obvious which analysis items that have influence on the system functions.

For complex systems with an ample degree of redundancy or with buffers, we may need a formal approach to identify the functional significant items.

If failure rates and other necessary input data are available for the various analysis items, it is usually a straightforward task to calculate the relative importance of the various analysis items based on a fault tree model or a reliability block diagram. A number of importance measures are discussed by Rausand and Høyland (2003). In a Monte Carlo model it is also rather straightforward to rank the various analysis items according to criticality.

The main reason for performing this task is to screen out items that are more or less irrelevant for the main system functions, i.e. in order not to waste time and money analyzing irrelevant items.

In addition to the FSIs, we should also identify items with high failure rate, high repair costs, low maintainability, long lead time for spare parts, or items requiring external maintenance personnel. These analysis items are denoted maintenance cost significant items (MCSI).

The sum of the functional significant items and the maintenance cost significant items are denoted maintenance significant items (MSI).

Some authors, e.g. Smith (1993), claim that such a screening of critical items should not be done, others e.g. Paglia *et al.* (1991) claim that the selection of critical items is very important in order not to waste time and money. We tend to agree with both. In some cases it may be beneficial to focus on critical items, in other cases we should analyse all items.

In the RCM project for the Norwegian Railway Administration the use of generic RCM analyses made it possible to analyse all identified MSIs. Thus this step tend to be less critical if a generic approach is taken.

In the FMECA analysis of Step 6, each of the MSIs will be analysed to identify their possible impact upon failure on the four consequence classes: (S) safety of personnel, (E) environmental impact, (A) production availability (punctuality), and (C) economic losses. This analysis is partly inductive and will focus on both local and system level effects.

10.5 Step 5: Data collection and analysis

The purpose of this step is to establish a basis for both the qualitative analysis (relevant failure modes and failure causes), and the quantitative analysis (reliability parameters such as MTTF, PF intervals and so on). See Chapters 13 and 14 for elements of data collection and analysis

10.6 Step 6: Failure modes, effects and criticality analysis

The objective of this step is to identify the dominant failure modes of the MSIs identified during Step 4. The FMECA methodology is discussed in Chapter 15.

10.7 Step 7: Selection of Maintenance Actions

This phase is the most novel compared to other maintenance planning techniques. A decision logic is used to guide the analyst through a question–and–answer process. The input to the RCM decision logic is the dominant failure modes from the FMECA in Step 6. The main idea is for each dominant failure mode to decide whether a preventive maintenance task is suitable, or it will be best to let the item deliberately run to failure and afterwards carry out a corrective maintenance task. There are generally three reasons for doing a preventive maintenance task:

a) to prevent a failure
b) to detect the onset of a failure
c) to discover a hidden failure

Only the dominant failure modes are subjected to preventive maintenance. To obtain appropriate maintenance tasks, the failure causes or failure mechanisms should be considered. The idea of performing a maintenance task is to prevent a failure mechanism to cause a failure. Hence, the failure mechanisms behind each of the dominant failure modes should be entered into the RCM decision logic to decide which of the following basic maintenance tasks that is applicable:

1. Continuous on–condition task (CCT)
2. Scheduled on–condition task (SCT)
3. Scheduled overhaul (SOH)
4. Scheduled replacement (SRP)
5. Scheduled function test (SFT)
6. Run to failure (RTF)

Continuous on–condition task (CCT) is a continuous monitoring of an item to find any potential failures. An on–condition task is applicable only if it is possible to detect reduced failure resistance for a specific failure mode from the measurement of some quantity.

Example:
A distance gauge on the turnout might be used to measure the distance between the switch point and stock rail to detect that the 3mm limit will be reached. At a predefined level (i.e.

2.7 mm), the system alerts the maintenance crew, which carry out an appropriate maintenance action.

Scheduled on–condition task (SCT) is a scheduled inspection of an item at regular intervals to find any potential failures. There are three criteria that must be met for an on–condition task to be applicable:

1. It must be possible to detect reduced failure resistance for a specific failure mode.
2. It must be possible to define a potential failure condition that can be detected by an explicit task.
3. There must be a reasonable consistent age interval between the time of potential failure and the time of failure.

Examples:
A manual inspection every second month will reveal whether the "3 mm limit" is soon being reached. Appropriate maintenance action can be issued. Ultrasonic inspection of rails every year to detect cracks in the rails.

There are two disadvantage of a scheduled versus a continuous on-condition task:

- The man-hour cost of inspection is often larger than the cost of installing the sensor
- Since the scheduled inspection is carried out at fixed points of time, one might "miss" situations where the degradation is faster than anticipated.

An advantage of a scheduled on-condition task is that the human operator is then able to "sense" information that a physical sensor will not be able to detect. This means that traditional "Walk around checks" should not be totally skipped even if sensors are installed.

Condition monitoring is discussed in Nowlan & Heap (1978), and statistical models are presented in e.g. Aven (1992) and Valdez-Flores & Feldman (1989).

Scheduled overhaul (SOH) is a scheduled overhaul of an item at or before some specified age limit, and is often called "hard time maintenance".

An overhaul task can be considered applicable to an item only if the following criteria are met (Nowlan & Heap 1978):

1. There must be an identifiable age at which the item shows a rapid increase in the item's failure rate function.
2. A large proportion of the units must survive to that age.
3. It must be possible to restore the original failure resistance of the item by reworking it.

Examples:
Rehabilitation of wooden sleepers borings every three year. Lubrication of the char-/slideplate every three day. Cleaning every month.

Scheduled replacement (SRP) is scheduled discard of an item (or one of its parts) at or before some specified age limit. A scheduled replacement task is applicable only under the following circumstances (Nowlan & Heap 1978):

1. The item must be subject to a critical failure.
2. Test data must show that no failures are expected to occur below the specified life limit.

3. The item must be subject to a failure that has major economic (but not safety) consequences.
4. There must be an identifiable age at which the item shows a rapid increase in the failure rate function.
5. A large proportion of the units must survive to that age.

Example:
Replacement of the motor every one year The motor is then either overhauled to “a god as new” condition, or replaced in the maintenance depot.

Scheduled function test (SFT) is a scheduled inspection of a hidden function to identify any failure. A scheduled function test task is applicable to an item under the following conditions (Nowlan & Heap 1978):

1. The item must be subject to a functional failure that is not evident to the operating crew during the performance of normal duties.
2. The item must be one for which no other type of task is applicable and effective.

Example:
Sighting or hammer blow every year to detect loose lockspikes fastening chars/baseplates on wooden sleepers.

Run to failure (RTF) is a deliberate decision to run to failure because the other tasks are not possible or the economics are less favourable.

In many situations one maintenance task may prevent several failure mechanisms. Hence in some situations it is better to put failure modes rather than failure mechanisms into the RCM decision logic.

Note also that if a failure cause for a dominant failure mode corresponds to a supporting equipment, the supporting equipment should be defined as the “item” to be entered into the RCM decision logic.

The criteria given for using the various tasks should only be considered as guidelines for selecting an appropriate task. A task might be found appropriate even if some of the criteria are not fulfilled.

The RCM decision logic is shown in Figure 45. Note that this logic is much simpler than those found in standard RCM references, e.g. Moubray (1991). It should be emphasized that such a logic can never cover all situations. For example in the situation of a hidden function with ageing failures, a combination of scheduled replacements and function tests is required.

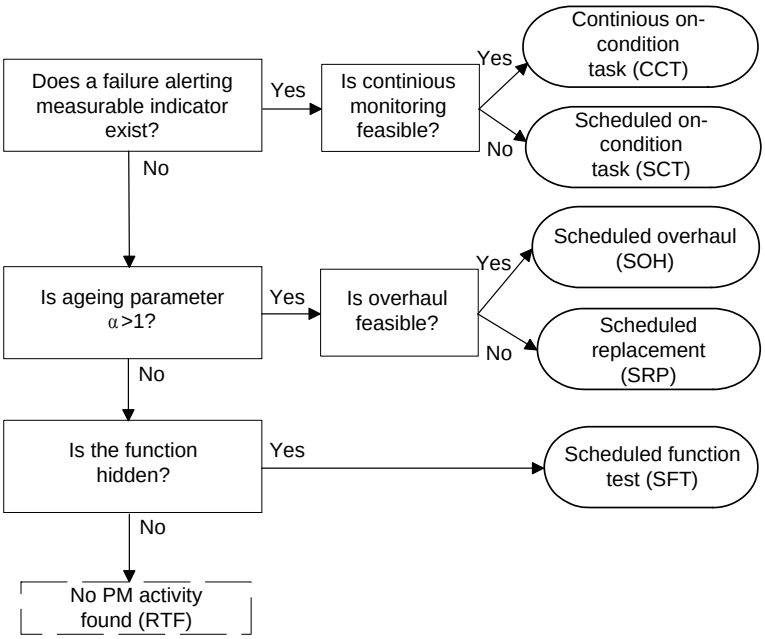


Figure 45 Maintenance Task Assignment/Decision logic

10.8 Step 8: Determination of Maintenance Intervals

Usually formalised methods for optimisation of maintenance interval is not a part of the RCM analysis. In order to optimise maintenance intervals we need to structure the analysis in such a way that it fits into the maintenance optimisation models that exists. See Chapter 11 for a discussion of determination of maintenance intervals using optimisation models.

10.9 Step 9: Preventive maintenance comparison analysis

Two overriding criteria for selecting maintenance tasks are used in RCM. Each task selected must meet two requirements:

- It must be applicable
- It must be effective

Applicability: meaning that the task is applicable in relation to our reliability knowledge and in relation to the consequences of failure. If a task is found based on the preceding analysis, it should satisfy the Applicability criterion.

A PM task will be applicable if it can eliminate a failure, or at least reduce the probability of occurrence to an acceptable level (Hoch 1990) - or reduce the impact of failures!

Cost-effectiveness: meaning that the task does not cost more than the failure(s) it is going to prevent.

The PM task's effectiveness is a measure of how well it accomplishes that purpose and if it is worth doing. Clearly, when evaluating the effectiveness of a task, we are balancing the "cost" of "performing the maintenance with the cost of not performing it. In this context, we may refer to the cost as follows (Hoch 1990):

1. The "cost" of a PM task may include:
   - the risk of maintenance personnel error, e.g. "maintenance introduced failures"

- the risk of increasing the effect of a failure of another component while the one is out of service
- the use and cost of physical resources
- the unavailability of physical resources elsewhere while in use on this task
- production unavailability during maintenance
- unavailability of protective functions during maintenance of these
- "The more maintenance you do the more risk you will expose your maintenance personnel to"

2. On the other hand, the "cost" of a failure may include:
   - the consequences of the failure should it occur (i.e. loss of production, possible violation of laws or regulations, reduction in plant or personnel safety, or damage to other equipment)
   - the consequences of not performing the PM task even if a failure does not occur (i.e., loss of warranty)
   - increased premiums for emergency repairs (such as overtime, expediting costs, or high replacement power cost).

10.10Step 10: Treatment of non-MSIs

In Step 4 critical items (MSIs) were selected for further analysis. A remaining question is what to do with the items which are not analysed. For plants already having a maintenance program it is reasonable to continue this program for the non-MSIs. If a maintenance program is not in effect, maintenance should be carried out according to vendor specifications if they exist, else no maintenance should be performed. See Paglia *et al* (1991) for further discussion.

10.11Step 11: Implementation

A necessary basis for implementing the result of the RCM analysis is that the organizational and technical maintenance support functions are available. A major issue is therefore to ensure the availability of the maintenance support functions. The maintenance actions are typically grouped into maintenance packages, each package describing what to do, and when to do it.

Many accidents are related to maintenance work. When implementing a maintenance program it is therefore of vital importance to consider the risk associated with the execution of the maintenance work. Checklists could be used to identify potential risk involved with maintenance work:

- Can maintenance people be injured during the maintenance work?
- Is work permit required for execution of the maintenance work?
- Are means taken to avoid problems related to re-routing, by-passing etc.?
- Can failures be introduced during maintenance work?
- etc.

Task analysis, see e.g. Kirwan & Ainsworth (1992) may be used to reveal the risk involved with each maintenance job. See Hoch (1990) for a further discussion on implementing the RCM analysis results.

10.12Step 12: In-service data collection and updating

As mentioned earlier, the reliability data we have access to at the outset of the analysis may be scarce, or even second to none. In our opinion, one of the most significant advantages of RCM is that we systematically analyze and document the basis for our initial decisions, and, hence, can better utilize operating experience to adjust that decision as operating experience data is collected. The full benefit of RCM is therefore only achieved when operation and maintenance experience is fed back into the analysis process.

The process of updating the analysis results is also important due to the fact that nothing remain constant, best seen considering the following arguments (Smith 1993):

- The system analysis process is not perfect and requires periodic adjustments.
- The plant itself is not a constant since design, equipment and operating procedures may change over time.
- Knowledge grows, both in terms of understanding how the plant equipment behaves and how technology can increase availability and reduce costs.

Reliability trends are often measured in terms of a non-constant ROCOF (rate of occurrence of failures), see e.g. Rausand and Høyland (2003). The ROCOF measures the probability of failure as a function of *calendar* time, or global time since the plant was put into operation. The ROCOF may change over time, but within one cycle the ROCOF is assumed to be constant. This means that analysis updates should be so frequent that the ROCOF is fairly constant within one period.

Opposite to the ROCOF, the *failure rate* or FOM, is measuring the probability of failure as a function of *local* time, i.e. the time elapsed since last repair/replacement. However, the FOM can not be considered constant, if so there is no rationale for performing scheduled replacement/repair.

The updating process should be concentrated on three major time perspectives (Sandtorv & Rausand 1991):

- Short term interval adjustments
- Medium term task evaluation
- Long term revision of the initial strategy

The short term update can be considered as a revision of previous analysis results. The input to such an analysis is updated reliability figures either due to more data, or updated data because of reliability trends. This analysis should not require much resources, as the framework for the analysis is already established. Only Step 5 and Step 8 in the RCM process will be affected by short term updates.

The medium term update will also review the basis for the selection of maintenance actions in Step 7. Analysis of maintenance experience may identify significant failure *causes* not considered in the initial analysis, requiring an updated FMECA analysis in Step 6. The medium term update therefore affects Step 5 to 8.

The long term revision will consider all steps in the analysis. It is not sufficient to consider only the system being analysed, it is required to consider the entire plant with it's relations to the outside world, e.g. contractual considerations, new laws regulating environmental protection etc.

10.13Generic and local RCM analysis

In principle, the RCM analysis should be conducted for *physical* units in an explicit operational context. This means that we for example conduct an RCM analysis for a given turnout at location *X* at line *Y.* For this turnout we identify all functions, failure modes etc. Then we propose a set of maintenance tasks, and finally chose the maintenance intervals based on the reliability performance parameters for that turnout, and the personnel and punctuality risk for that turnout. Now, there might be several hundreds of similar turnouts, but where both the reliability performance and the risk profile might vary, which again should ask for different maintenance intervals. The question is whether we need to repeat the entire RCM analysis for all the (similar) turnouts? The proposed answer to this question is to first conduct a *generic* RCM analysis, and then perform local adjustment to risk parameters. The following steps would then be required:

1. ***Conduct a generic RCM analysis for selected components***. In this analysis we use generic, or average values of reliability parameters, and consequences parameters describing safety and punctuality risk.
2. ***Generic RCM database***. The results from the generic RCM analysis is stored in a generic RCM database, i.e. generic analyses for selected equipment types. These types could be e.g. a turnout, a main signal, traction system, break system etc. In the first place we might restrict ourselves to consider a broad class of e.g. turnouts (different manufactures). In a later phase we might want to refine our analysis to also consider qualitative different turnouts (with different failure modes).
3. ***Selection of local analysis objects***. In the local analysis we work with a subset of the railway system. This could be for example one specific line, turnouts in the main track of one specific line, one specific train set, one specific train set operating on one specific line etc.
4. ***Find an appropriate generic RCM template***. For a local analysis object, we now recall the corresponding generic RCM analysis from the RCM database. We first verify that the generic RCM analysis object (template) is appropriate in terms of qualitative properties, i.e. the different functions, failure modes etc that are considered. At this point it might be necessary to add more failure modes, regard some failure modes etc. If this is the case, we add the "new" RCM object to the generic RCM database in order to make the generic RCM database more and more comprehensive.
5. ***Adjust parameters***. At the local level we identify differences from the generic parameters used in the generic RCM database. For example a specific line might have very old turnouts, and hence the MTTF is shorter than the average MTTF. At this step of the procedure we have to consider all parameters that are involved in the optimisation model (see Chapter 10).
6. ***Re-run the optimisation procedure.*** Based on the new "local" parameters we will re-run the optimisation procedure to adjust maintenance intervals taking local differences into account. To carry out this process we need a computerised tool to streamline the work.
7. ***Document the results.*** The results from the local analysis is stored in a local RCM database. This is a database where only the adjustment factors are documented, for example for turnouts A, B, C and D on line Y the MTTF is 30% higher than the average. Hence the maintenance interval is also reduced accordingly.

10.14Risk based inspection

Risk based inspection (RBI) is an approach to establish an inspection strategy for a plant. The methodology is in many aspects similar to the RCM approach. Some main differences between RCM and RBI are:

- RCM is a general method that could be applied a wide range of applications, whereas RBI is a tailor-made method which only applies typically for structural elements where the degradation could be measured, i.e. by means of inspection.
- RBI manuals usually cover a wide range of inspection methods and a discussion of the applicability of the various methods in different situations.
- The RBI method is much more integrated with the risk management system than usually is the case for RCM. This means that the safety implication of failures are more explicitly treated, and risk is often quantified on a detailed level, and compared with the overall risk acceptance criteria for the plant.

Some references to RBI are:

- The DNV recommended practice, Risk Based Inspection of Offshore Topside Static Mechanical Equipment. (DNV-RP-G101, see http://exchange.dnv.com).
- Best practice for risk based inspection as a part of plant integrity management (Wintle *et.al* 2001).
- API Recommended Practice 580, Risk-Based Inspection. (http://www.techstreet.com/cgi-bin/detail?product_id=959810)

Wintle *et.al* (2001) proposes the following steps in a process diagram for plant integrity management by RBI:

1. Assess the requirements for integrity management and risk based inspection
2. Define the systems, the boundaries of systems, and the equipment requiring integrity management
3. Specify the integrity management team and responsibilities
4. Assemble plant database
5. Analyse accident scenarios, deterioration mechanisms, and assess and rank risks and uncertainties
6. Develop inspection plan within integrity management strategy
7. Achieve effective and reliable examination and results
8. Assess examination results and fitness-for-service

9a. Update plant database and risk analysis, review inspection plan and set maximum intervals to next examination

9b. Repair, modify, change operating conditions

10. Audit and review integrity management process

11. SIMPLIFIED RISK MODELLING AND OPTIMISING

This chapter is primarily intended for risk modelling when optimising maintenance intervals as a part of an RCM analysis. When structuring of the risk picture we have aimed at establishing a model that could be reflected in the columns of the FMECA, see Section 15.4.

In order to optimise maintenance we need a risk model on a format that allows us to predict the risk level as a function of the maintenance level. Such a model has two major part:

- A model that shows the relation between maintenance effort and component performance
- A model that shows the relation between component performance and system risk

The component model will typically involve the calculation of the "effective" failure rate as a function of the maintenance interval τ. The system model will be a combination of fault tree analysis, event tree analysis, Markov models and so forth. If such models have been developed for the system that is being analysed with respect to maintenance optimisation we may use these models. However, often such models do not exist and it will require too much effort to develop them. If this is the case we would rather develop a much simpler system risk model. We will now present such a simplified risk model, and discuss how we could use this model for optimising preventive maintenance. We will show the "safety" part, and the "punctuality" part of the model. Other dimensions could also be included if necessary.

11.1 Simplified safety modelling

The safety model is shown in Figure 46.

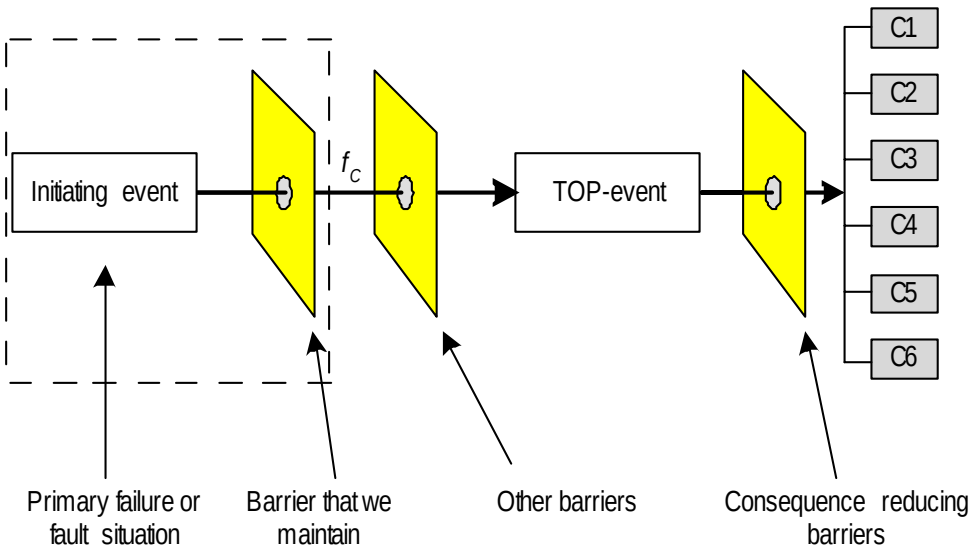


Figure 46 Barrier model for safety

In the dotted rectangle to the left we have an "initiating event" and a "barrier". To describe the content of this rectangle explicit we need reliability parameters as MTTF, ageing parameter, PF-interval etc described in the FMECA analysis, see Section 15.4. There are basically three situations that are considered:

1. There is a failure or a fault situation that is not related to the component we are analysing with respect to maintenance. For example we are analysing the ATP (Automatic Train Protection) on the train. In this situation the initiating event could be "locomotive driver does not comply with signalling", and thus the ATP is a barrier against this initiating event. In this situation the function of the ATP is typically a *hidden function*.
2. There is a potential failure in the component that are being analysed, and maintenance is a barrier against this failure. For example a crack is initiated in the rail, or in an axel (initiating event), and ultrasonic inspection is a maintenance activity to reveal the crack, and prevent a serious incident.
3. The initiating event is a component failure, and preventive maintenance is carried out to reduce the likelihood of this failure. In this situation the "initiating event" and the first "barrier" in Figure 46 merges to one "element". An example is ageing failure of a light bulb. The likelihood of such a failure will however be reduced if the light bulb is periodically replaced with a new one before the ageing effect becomes dominant.

The "other barriers" represents other barriers that could prevent the component failure from developing further to a critical event, or the TOP-event. For example "track circuit detection" is a barrier against rail breakage, because the track circuit could detect a broken rail. In the FMECA form described in Section 15.4 the "other barriers" are described both qualitatively, and quantitatively ($P_{TE\text{-}S}$)

The TOP-event is in this context the accidental event. Within railway application it is common to define the following seven TOP events:

- Derailment
- Collision train-train
- Collision train-object
- Fire
- Persons injured or killed in or at the track
- Persons injured or killed at level crossings
- Passengers injured or killed at platforms

If the TOP-event occurs there could also be consequence reducing barriers. For example the use of guide rails will usually have a very good impact on derailments.

In Figure 46 we have finally indicated that the outcome of the TOP event could be one of six consequence classes:

C_1: Minor injury
C_2: Medical treatment
C_3: Permanent injury
C_4: 1 fatality
C_5: 2-10 fatalities
C_6: >10 fatalities

Figure 46 is a simplified model for the risk picture related to the component that is being analysed. In order to quantify the risk we need the following quantities:

f_I = the frequency of the initiating event
Q_M = the probability that the maintained barrier does not function as intended
$P_{TE\text{-}S}$ = probability that the other barriers against the TOP-event all fails
P_{Cj} = probability that the TOP-event results in consequence C_j, j = 1,..,6

The frequency of the consequence classes C_j are now given by:

$$F_j = f_I \times Q_M \times P_{TE\text{-}S} \times P_{Cj} \quad (82)$$

We will later on indicate how we may model equation (82) as a function of the maintenance interval, τ.

In some situation we also assign a cost, or a PLL (Potential Loss of Life) contribution to the various cost elements. Proposed values are given in Table 5. Please see discussion in e.g. Vatn (1998) regarding what it means to assign monetary values to safety. The cost figures below have been adopted by the Norwegian Railway Administration.

Table 5 PLL-contribution and Cost contribution to the consequence classes

Consequence	PLL_j = PLL-contribution	SC_j = Cost (NOK)	Cost (Euro)
C_1: Minor injury	0.01	15 000	2 000
C_2: Medical treatment	0.05	250 000	30 000
C_3: Permanent injury	0.1	2 500 000	300 000
C_4: 1 fatality	0.7	13 000 000	1 600 000
C_5: 2-10 fatalities	4.5	100 000 000	13 000 000
C_6: >10 fatalities	30	1 300 000 000	160 000 000

The total PLL contribution related to the component being analysed is then:

$$\text{PLL} = f_I \times Q_M \times P_{TE\text{-}S} \times \sum_{j=1:6}(\text{PC}_j \times \text{PLL}_j) \quad (83)$$

And the total cost contribution related to the component is

$$C_S = f_I \times Q_M \times P_{TE\text{-}S} \times \sum_{j=1:6}(\text{PC}_j \times \text{SC}_j) \quad (84)$$

Table 6 Generic probabilities, PC_j, of consequence class C_i for the different TOP events

TOP event	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6
Derailment	0.1	0.1	0.1	0.1	0.05	0.01
Collision train-train	0.02	0.03	0.05	0.5	0.3	0.1
Collision train-object	0.1	0.2	0.3	0.15	0.01	0.001
Fire	0.1	0.2	0.2	0.1	0.02	0.005
Passengers injured or killed at platforms	0.3	0.3	0.2	0.05	0.01	0.001
Persons injured or killed at level crossings	0.1	0.2	0.3	0.3	0.09	0.01
Persons injured or killed in or at the track	0.2	0.2	0.2	0.3	0.1	0.0001

Note that we in the FM ECA analysis could have an automatic procedure that calculates the PLL contribution, and the safety cost contribution based on the reliability parameters, and the type of TOP event, see also Section 15.4.

Exercise 15
Consider a situation where a (hidden) safety function is demanded with frequency $f_I = 10^{-3}$ per year. The safety function is assumed to have exponentially distributed time to failure with MTTF = 2 years. If the safety function is demanded, and it fails, then the TOP event (derailment) will occur with a probability $P_{TE\text{-}S} = 0.05$. Assume the safety function is tested twice a year. Find the frequency F_j for each consequence class by using Table 6. □

Exercise 16
Consider exercise **15** and calculate the PLL and cost contributions in this situation. What will be the economical gain in terms of reduced safety costs if the test is conducted 4 times a year. □

11.2 Punctuality modelling

The risk model for punctuality is very similar to the risk model for safety and is shown in Figure 47.

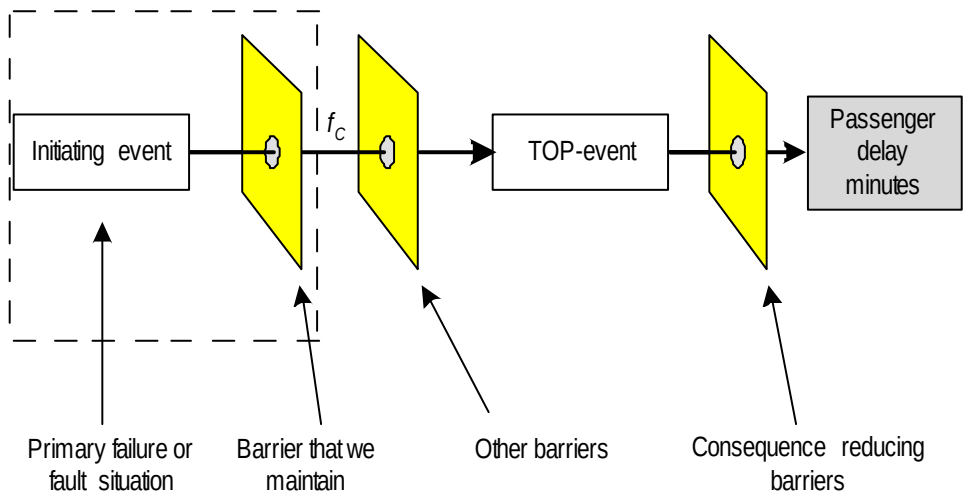


Figure 47 Risk model for punctuality

From the left, the model is identical to the safety model up to the "TOP" event, except for notation where we used $P_{TE\text{-}P}$ for TOP-event (barrier) probability for punctuality. The following TOP events for punctuality is proposed:

- Full stop (Infrastructure)
- Slow speed (Infrastructure)
- Manual train operation – line block (Infrastructure)
- Manual train operation – station (Infrastructure)
- Full stop – First line maintenance (Rolling stock)
- Full stop – Depot maintenance (Rolling stock)
- ATP failure–80 km/h (Rolling stock)
- Slow speed –40 km/h (Rolling stock)

(list to be completed…)

The relation between the TOP-event and "Passenger delay minutes" is generally very complex. It is far outside the scope of this presentation to present a mathematical model for this relation. The following factors should at least be taken into account:

Table 7 Factors influencing passenger delay minutes

Factor	Notation	Unit	Comment/values
Repair time	MTTR	Minutes	
Availability of rescue train	ART		1 = Good, 2 = Bad, 5 =Very bad
Mobilisation time	MoT	Minutes	
Single track/double track	SDT		1 = Double track, 2 = Single track
Train density	TrD	Trains/hour	
Length of line blocks	LLB	km	
Line speed	LSp	Km/h	
Passengers per train	PPT	#	
TOP event specific factor	TEF		To be defined!

A very simple model for passenger delay minutes (PDM) is now:

$$PMD = (MTTR+MoT) \times ART \times PPT \times LSp/100 \times (1+LLB/10) \times (1 + TrD/4) \times SDT \times TEF(85)$$

The punctuality cost could then be found as

$$C_P = f_I \times Q_M \times P_{TE\text{-}P} \times PMD \quad (86)$$

Exercise 17
Consider a situation with a engine breakdown that requires a rescue train. Calculates passenger delay minutes (PDM) when MTTR = 1 hour, MoT = 2 hours, ART = 2 (bad), SDT = 1 (double track), TrD = 10, LLB = 5, LSp = 160, PPT = 250. TEF = 1. □

Exercise 18
How well is the punctuality model calibrated in relation to you understanding of passenger delay minutes in the situation described in exercise **17**? Propose a new value for TEF in this situation based on you understanding. □

Table 8 Punctuality cost per passenger minute delay

Situation	PMD cost (NOK)	PMD cost (Euro)
High number of business travellers	5	0.6
Average number of business travellers	3	0.4
Low number of business travellers	1	0.13

Exercise 19
Consider the situation in exercise **17**, and assume that there is a high number of business travellers on the line where the breakdown most likely will occur. What is the punctuality cost of an engine failure if the TOP event occurs. □

11.3 Modelling the effect of maintenance on component level

In order to finalise the optimisation model we need to assess the component performance (the inside the dotted rectangle of Figure 46). The aim is to find the frequency of "failures", f_C of the dotted rectangle of Figure 46, and we start by defining:

MTTF	Mean Time To Failure without maintenance
α	Ageing parameter. Typically $\alpha = 2$ corresponds to weak, $\alpha = 3$ to medium, and $\alpha = 4$ to strong ageing.
E_{PF}	Expected value for the P-F interval
SD_{PF}	Standard deviation for the P-F interval. If this information is not available, $SD_{\text{P-F}} = 0.5 \times E_{\text{P-F}}$
P_I	Probability that an inspection will reveal a potential failure. If P_I could not be quantified, use as a rule of thumb $P_I = 0.9$ for good detection probability, $P_I = 0.7$ for medium detection probability, and $P_I = 0.2$ for low detection probability.
$\lambda_A(\tau)$	Effective failure rate as a function of the maintenance interval. See equation (70) page 73 for exact formulas. For approximate formulas use equation (58) page 66. Notate that to calculate $\lambda_A(\tau)$ we also need values for the parameters in the Weibull distribution, i.e. the ageing parameter α, and MTTF.
$Q_{PF}(\tau)$	The probability that the inspection strategy will succeed in revealing an initiated failure progression (i.e. a crack) in due time. $Q_{PF}(\tau)$ could be found by reading from Figure 31 page 64. Note that we need values for the relevant parameters, that is E_{PF}, SD_{PF} and P_I.
f_P	Frequency of "potential failures", i.e. the number of "P"s in the "PF-interval" per time unit. $f_P = 1/(\text{MTTF} + E_{PF})$.
f_D	Demand rate for which the hidden function is demanded. For example if the maintenance object is a stroke detector, then f_D is the frequency of train with bad wheels. For a fire detector, f_D is the frequency of fires etc,
τ_A	Interval for preventive replacement/overhaul for ageing components
τ_{PF}	Interval for condition monitoring (PF situation, Failure progression)
τ_{FT}	Interval for functional test (hidden function)

Table 9 now shows the relation between the maintenance interval (τ_A, τ_{PF} and/or τ_{FT}) and the value of f_C.

Table 9 f_C as a function of maintenance interval

Operational situation→ ↓ Failure progression	Evident function/ continuous demand	Hidden function/ spurious demand
Obs. failure progression (PF int.)	$f_I = f_P$ $Q_M = Q_{PF}(\tau_{PF})$ $f_C = f_P \times Q_{PF}(\tau_{PF})$	$f_I = f_D$ $Q_M = Q_{PF}(\tau_{PF}) \times E_{PF} \times f_P$ $f_C = Q_{PF}(\tau_{PF}) \times f_D \times E_{PF} \times f_P$
Ageing	$f_I = 1/MTTF$ $f_C = \lambda_A(\tau_A)$	$f_I = f_D$ $Q_M = \lambda_A(\tau_A) \times \tau_{FT}/2$ $f_C = f_D \times \lambda_A(\tau_A) \times \tau_{FT}/2$
Random	$f_I = 1/\text{MTTF}$ $f_C = 1/\text{MTTF}$	$f_I = f_D$ $Q_M = \tau_{FT}/(2\text{MTTF})$ $f_C = f_D \times \tau_{FT}/(2\text{MTTF})$

Exercise 20
Consider the situation in exercise **17**, and now assume the following simplified model for the engine: MTTF (without maintenance=engine revision) is 5 years. We assume medium ageing (α = 3). Further, if the engine fails, we assume that the TOP event occurs with probability $P_{TE\text{-}P}$ = 0.3. Find the punctuality cost as a function of maintenance interval, i.e. let τ_A be the interval length for revision. Hint: Use the "Ageing" for "continuous demand". Calculate the cost for τ_A = 1, 2 and 3 years. □

11.4 Optimisation of preventive maintenance

In this section we presented the basic models that are required to optimise maintenance intervals. We have:

- Established models that could be used to find the relation between maintenance intervals and the component failure frequency, f_C. (Table 9)
- A risk model for safety (Figure 46) and for punctuality (Figure 47), and formulas for safety and punctuality costs.

By combining these results we may in principle obtain the total safety and punctuality cost. If we now also add preventive and corrective maintenance cost, we could obtain the total cost per unit time by:

$$C(\tau) = C_S(\tau) + C_P(\tau) + C_{PM}(\tau) + C_{CM}(\tau) \quad (87)$$

Where $C_S(\tau)$ and $C_P(\tau)$ are found by equations (84) and (86) respectively. Further

$$C_{PM}(\tau) = \mathrm{PM}_{Cost}/\tau \quad (88)$$

Where PM_{Cost} is the cost per preventive maintenance activity. Further if CM_{Cost} is the cost of a corrective maintenance activity, we have

$$C_{CM}(\tau) = \mathrm{CM}_{Cost} \times f_C \quad (89)$$

To find the optimum maintenance interval we could then in principle calculate $C(\tau)$ from equation (1) for various values of the maintenance interval, τ, and then chose the τ-value that minimises $C(\tau)$.

Exercise 21
Use the Excel sheet to optimise maintenance interval in a situation you are familiar with. Include both safety and punctuality cost. □

11.5 Grouping of maintenance action

In Section 11.4 we have indicated a method for choosing a maintenance interval that minimises the total cost per unit time. In this approach we have been considering one component, or failure mode, at a time. In real life we would, however, consider several maintenance action in one "work package". For example if we preventively will replace a light bulb in a departure light signal, we would also consider other maintenance activities, such as cleaning the lenses, controlling the transformer etc. To model such a situation the complexity of the problem increases dramatically. In a situation where we take for granted which activities that should be grouped it is rather simple to carry out the optimisation. However, if we also want do determine an optimal grouping strategy, the problem is far outside the scope for this presentation. See e.g. Wildeman (1996) for an introduction to this

topic. In the following we will discuss some basic elements of modelling the cost structure when the grouping is *given*.

As a starting pint we consider the cost per unit time in Equation (1). Now, for simplicity, assume that we have two components A and B, and that the optimum maintenance interval for each of them using Equation (1) is in the same order of magnitude. We would expect to achieve some cost savings due to sharing set-up costs if we combine the activities, which could result in a reduction of the optimal maintenance interval. We will first investigate the PM cost. For each of the component we let $PM_{Cost,A}$ and $PM_{Cost,A}$ denote the cost if PM activity is carried out separately for the two components A and B respectively. Now, assume that the PM_{Cost} could be split into a common set-up cost when maintenance of A and B are combined. We denot tis cost $PM_{Cost,S}$. The remaining part of the PM cost for each of the components is assumed to be $PM_{Cost,A}$ - $PM_{Cost,S}$ and $PM_{Cost,B}$ - $PM_{Cost,S}$ for component A and B respectively. Note that in railway maintenance the set-up cost will often dominate the cost per component, at least for infrastructure components where traveling to the cite and rigging is the main contributor to the cost. We now have the preventive maintenance cost per unit time for the two components:

$$C_{PM}(\tau) = (PM_{Cost,A} + PM_{Cost,B} - PM_{Cost,S}) / \tau \tag{90}$$

If we treat the CM cost, it is not reasonable to have any synergy effects here, hence

$$C_{CM}(\tau) = CM_{Cost,A} \times f_{C,A} + CM_{Cost,B} \times f_{C,B} \tag{91}$$

Where index A and B refer to the two components. Note that the frequency f_C is affected by the maintenance interval through the relations given in Table 9.

Now, let us consider the “system” cost, i.e. the safety cost and the punctuality cost. As a first approximation, we could treat these costs independently of each other for component A and B. For example, when we treat component A we calculate $f_{C,A}$, then find the probability that a failure in component A will cause the TOP event, and multiply these figures with the expected cost for the TOP event. We may then do the same for component B, and add the contribution for the two components, i.e. for safety:

$$C_S(\tau) = C_{S,A}(\tau) + C_{S,B}(\tau) \tag{92}$$

In principle, however, we should also investigate if one of the components A or B is a barrier against a failure of the other. For example, a reflex brand on a signalling pole is a barrier against a light bulb failure. In this situation we then need an explicit modelling of the interaction between these two “barriers”.

When the cost elements are found in this manner, we sum up all cost elements and choose the maintenance interval that minimises the total cost per unit time. The method outlined here could easily be extended to deal with more than two components.

12. OPTIMISATION OF RENEWAL

In this approach the objective is to establish a sound basis for the optimisation of maintenance and renewal. Different "headings" are used for such analysis, e.g. LCC analysis, Cost/Benefit analysis and NPV (Net Present Value) analysis. In all these situations the idea is to choose maintenance activities in time and space such that costs are minimised in the long run. The basic situation is that the railway infrastructure is deteriorating as a function of time and operational load. This is why the right part of the bath tube curve in Figure 2 is increasing. This deterioration could be transformed into cost functions, and when the costs become very large it might be beneficial to maintenance or renew the infrastructure. In the following we introduce the notation $c(t)$ for the costs as a function of time. In $c(t)$ we include in principal costs related to *i*) punctuality loss, *ii*) accident costs, and *iii*) extra maintenance and operation cost due to reduced track quality. By a maintenance or renewal action we typically reset the function $c(t)$, either to zero, or at least a level significantly below the current value. Thus, the operating costs will be reduced in the future if we are willing to invest in a maintenance or renewal project.

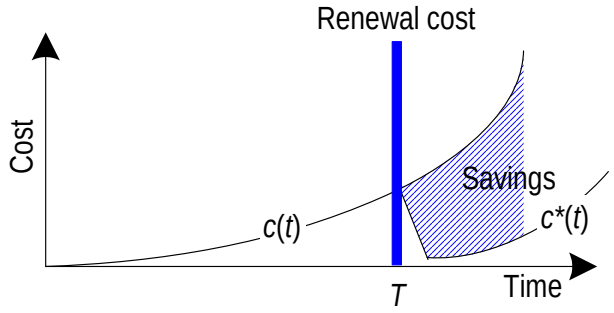


Figure 48 Cost savings

Figure 48 shows the savings in operational costs, $c(t)$ - $c^*(t)$, if we perform maintenance or renewal at time T. In addition to the savings in operational costs, we will also often achieve savings due to an increased "residual life time".

Special attention will be paid to projects that aim at extending the life length of a railway system. A typical example is rail grinding for extending the life length of the rail, but also for the fastenings, sleepers and the ballast. Figure 49 shows how a smart activity ( ) may suppress the increase in $c(t)$ and thereby extend the point of time before the cost explodes and a renewal is necessary.

12.1 Model input

In this section the basic input to the model is described. The description of each maintenance or renewal project could be stored in an MS ACCESS database.

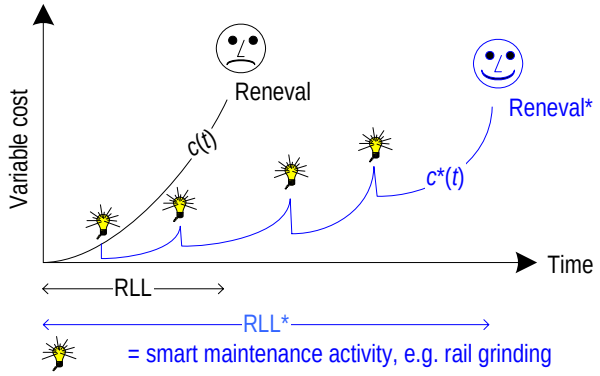


Figure 49 Life length extension

12.1.1Qualitative information

The situation leading up to each proposed project is described. This is typically information from measurements and analysis of track quality, trends etc.

12.1.2Safety related information

A general risk model has been derived where important risk influencing factors (RIF) has been identified. The RIFs relates both to the accident frequency such as number of cracks in the rails, but also to the accident consequences such as speed, terrain description etc. To describe the risk picture in a consistent manner, the user only has to enter the states or values related to the various RIFs. Then the program calculates the actual risk. In addition to the current value of the risk, also the future increase is described corresponding to the two cost curves $c(t)$ and $c^*(t)$ in Figure 48. Different functional forms could be entered, e.g. linear, exponential etc.

12.1.3Punctuality information

The basic punctuality information entered is the ordinary speed for the line, and any speed reductions due to the degradation the project is intended to fight against. The program then calculates the corresponding increase in travelling time. Very often such delays cause cascading effects in a tight network. Cascading effects could therefor also be entered. The user may also enter trend information.

12.1.4Maintenance and operating costs

The degradation of the permanent way will very often require extra maintenance and operating costs. Examples of such costs are extra runs of the measurement car, extra line inspections, use of alternative transportation such as busses, shorter lifetime of influenced components etc.

12.1.5 Residual life length

To be able to calculate the economical gain due to increased life lengths it is required to described the residual life length both if the proposed project is executed, e.g. RLL*, and if the project is not executed, RLL.

12.1.6 Project costs

The project costs are entered for each year in the project period.

12.1.7 Cost parameters

A set of general cost parameters is common for all projects. These are:

- The interest rent which is set to r = 4%.
- Monetary values for safety consequence classes as given in Table 10.
- Cost per minute kiloton freight delay =160 €.
- Cost per passenger minute delay = 0.4 €. A train with 250 passengers then gives 100 € per minute delay.

Table 10 Monetary values in € for each safety consequence class

Safety consequence		Monetary value
C_1	Minor injury	2 000
C_2	Medical treatment	33 000
C_3	Serious injury	330 000
C_4	1 fatality	1.7 millions
C_5	2-10 fatalities	11 millions
C_6	> 10 fatalities	175 millions

12.2 LCC calculation considerations

To calculate the various LCC contributions we need to consider three different aspects:

- Change in variable costs, $c(t)$.
- The effect of extending the life length.
- The project costs.

12.2.1 Change in variable costs

The variable cost contribution from the dimension safety; punctuality and maintenance & operation could be treated similarly from a methodical point of view. We now let $c(t)$ denote the variable cost if the project is not executed, and similarly $c^*(t)$ if the project is run. See Figure 48 for an illustration. The LCC contribution from change of e.g. safety could then be found by:

$$\Delta \mathrm{LCC}_S = \sum_{t=0}^{N} \left[c(t) - c^*(t) \right] \times (1+r)^{-t} \tag{93}$$

where r is the discounting factor, and N is the calculating period. We could either set N to a fixed value, e.g. 3 years, or we could set N to the residual life length, RLL if nothing is done.

Similarly we obtain the change in punctuality costs, ΔLCC_P and the change in maintenance and operational costs, $\Delta LCC_{M\&O}$.

12.2.2The effect of extending the life length.

To motivate for the calculation we show a principal sketch of the need for renewal both if or if not the proposed project is executed.

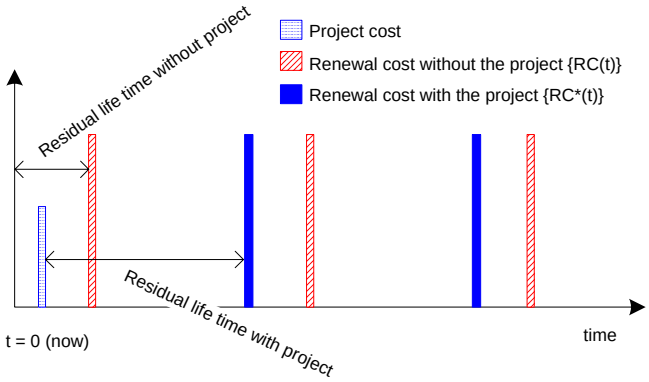


Figure 50 Renewals if and if not the project is executed

We now let:

- $\{RC(t)\}$ = Portfolio cost of renewals without the project
- $\{RC^*(t)\}$ = Portfolio costs of renewals with the project
- $\{T\}$ = Set of renewal times without the project
- $\{T^*\}$ = Set of renewal times with the project

The cost contribution related to increased residual life time could now be found by:

$$\Delta LCC_{RLT} = \sum_{t\in\{T\}} RC(t)\times(1+r)^{-t} - \sum_{t\in\{T^*\}} RC^*(t)\times(1+r)^{-t} \quad (94)$$

12.2.3The project costs

The LCC contribution from the project cost, LCC_I:, is the net present value of the project cost in the project period,

12.2.4Total LCC contribution

The total gain in terms of life cycle costs could then be found by:

$$\Delta LCC = LCC_I + \Delta LCC_S + \Delta LCC_P + \Delta LCC_{M\&O} + \Delta LCC_{RLT} \quad (95)$$

And the cost benefit ratio is:

$$\rho_{C/B} = \frac{\Delta LCC_S + \Delta LCC_P + \Delta LCC_{M\&O} + \Delta LCC_{RLT}}{LCC_I} \quad (96)$$

12.3 Example results

As a calculation example we will consider a rail-grinding project. Grooves and wave formations imply strong impact on the track and rolling stock due to increased dynamic loads and vibrations. This again gives shorter life length of the rails, but also to the sleepers, fastenings and ballast. Increased noise, energy consumption, and lower comfort could also be expected.

A 160-km section on the Rauma line in Norway has rail of age 40 to 50 years and rail grinding is recommended primarily to extend the life length of the rails.

12.3.1 Safety considerations

The derailment frequency due to rail breakages is estimated to 0.01 per year. For the most severe consequences we have the following distribution $P(C_4)$ = 13.5%, $P(C_5)$= 11% and $P(C_6)$ = 5% where the consequence classes are explained in Table 10. The material damages given a derailment is estimated to cost 1 300 000 €. Thus the yearly "safety costs" is found to be $0.01\times(0.135\times1.7 + 0.11\times11 + 0.05 \times 175 + 1.3)$ million €, which equals 110 000 €.

12.3.2 Punctuality costs

Due to a high number of cracks it is recommended to reduce the speed from 80 to 70 km/h for a section of 20 km. This corresponds to 2 minutes increase in travelling time. There are slightly more than 1000 passengers per week, thus the yearly delay time costs is in the order of 50 000 €. In addition there is also freight delay time costs in the order of 60 000 € per year.

12.3.3 Maintenance & operation costs

From different studies it is found that rail grinding every 40 megaton reduce the wear of other components corresponding to 8 € per meter. This corresponds to 500 000 € for the actual 160 km section.

12.3.4 Extended life length

By the rail grinding project it is assumed that the rails could be kept going for another 15 years, where as a rail renewal is expected after 5 years if the project is not run. The cost of new rails is in the order 250 € per meter. The life extension is estimated to 20% giving annual savings of approximately 50 € per meter, which gives 8 million € for the 160-km section. Also taking the discounting factor into account results in a saving of 11 million €.

12.3.5 Project costs

The cost of rail grinding is in the order of 8 € per meter, giving a total cost of 1.3 million €. In addition we have to expect a second grinding within 5 to 10 year, giving an additional contribution. The net present value of the grinding activity will then be 2.2 million €.

12.3.6 Cost benefit ratio

Summing up we find the following contribution to the change in LCC (million €):

ΔLCC_S = 0.5
ΔLCC_P = 0.6
$\Delta LCC_{M\&O}$ = 2.6
ΔLCC_{RLT} = 11
LCC_I = 2.2

This yields a cost benefit ratio of 6.6. This means that for every Euro put into rail grinding, the payback is almost 7 Euro.

13. SPECIFICATION OF A RAMS DATABASE

In this Chapter we give an outline of a proposed content of a database structure to be adopted in Railway Maintenance Management based on experience from the OREDA (Offshore REliability DAta) project. The database structure is based on a concept where failures and maintenance activities are linked to an inventory database. One inventory record corresponds to one physical equipment/component, for example one particular turnout.

For each inventory record there is a set of common variables/fields to enter, e.g. model, manufacturer and installation date. These common variables are listed in Table 11. In addition to the common variables there is also a set of equipment specific variables.

Failures and maintenance reports are linked to the inventory records. The set of common variables to enter for failures and maintenance reports are listed in Table 12 and Table 14 respectively.

Information about state variables (condition monitoring information) may also be entered into the RAMS database. For continuous measurements obtained by sensor technology, this information is linked directly to the inventory records, while information obtained during maintenance is linked to the maintenance records.

The relation between the various data tables is shown in Figure 53 page 117.

13.1 Relation to the OREDA project

The OREDA (Offshore Reliability Data) project has been running since the beginning of the eighties, and has been a joint effort between European oil companies. The guideline for collection of data within the OREDA project is now being implemented as an ISO standard (ISO 14224). The main principles for a railway RAMS database structure have been adapted from the ISO 14224, but modifications have been necessary. The following major changes compared to the ISO 14224 apply:

- Inclusion of condition monitoring (state information) data
- Failure mode identification at maintainable item level

13.2 Objectives

The main objective for a RAMS database is to facilitate systematic storage and retrieval of reliability and maintenance data. The information can be used both for strategic planning of maintenance and for reliability evaluation when approving new components. Some example of use of such a database is given below:

- Retrieval of qualitative information (“Upper ten lists”)
  - List of items frequently failing
  - List of frequently occurring failure causes
- Provide information on reliability parameters
  - Failure rates and life time *distributions*
  - Repair times
- Provide information regarding maintenance resources
  - Spare part consumption
  - Man-hours required (PM and CM)
- Provide condition monitoring information

- *Current* state of condition monitoring (CON) variables
- Correlation between failure probability and values of the CON variables
- Evolution of CON values as a function of time (how fast)

13.3 Equipment boundary and hierarchy

13.3.1 Boundary description

A clear boundary description is imperative for collecting, merging and analysing RAMS data from different industries, plants or sources. The merging and analysis will otherwise be based on incompatible data.

For each equipment class a boundary must be defined. The boundary defines what RAMS data are to be collected.

An example of a boundary diagram for a turnout is shown in Figure 51.

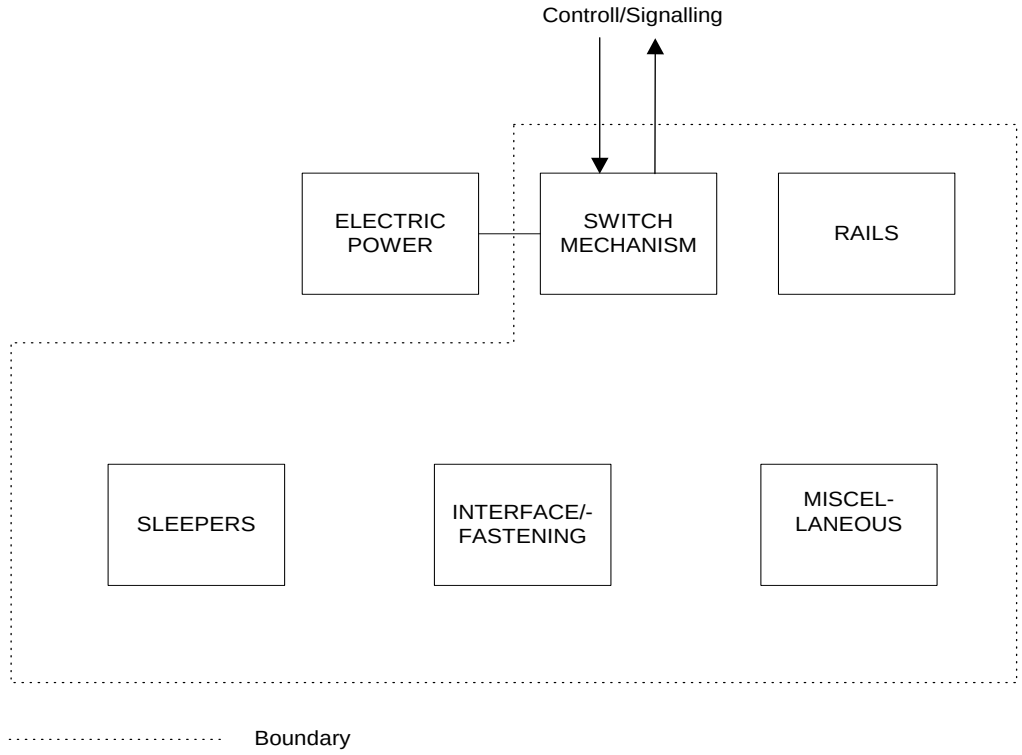


Figure 51 Example of boundary diagram (turnouts)

The boundary diagram shall show the subunits and the interfaces to the surroundings. Additional textual description shall, when needed for clarity, state in more detail what is to be considered inside and outside the boundaries.

13.3.2 Guidance for defining an equipment hierarchy

For the equipment it is recommended that a hierarchy is prepared. The highest level is the equipment unit class. The number of levels for subdivision will depend on the complexity of the equipment unit and the use of the data. Reliability data need to be related to a certain level

within the equipment hierarchy in order to be meaningful and comparable. For example, the reliability data “severity class” shall be related to the equipment unit while the failure cause shall be related to the lowest level in the equipment hierarchy.

A single instrument may need no further breakdown, while several levels are required for a compressor. For data used in availability analyses the reliability at the equipment unit level may be the only desirable data needed, while an RCM analysis will need data on failure mechanism at maintainable item level.

A subdivision into three levels for an equipment unit will normally be sufficient. An example is shown in Figure 52, viz. equipment unit, subunit and maintainable items.

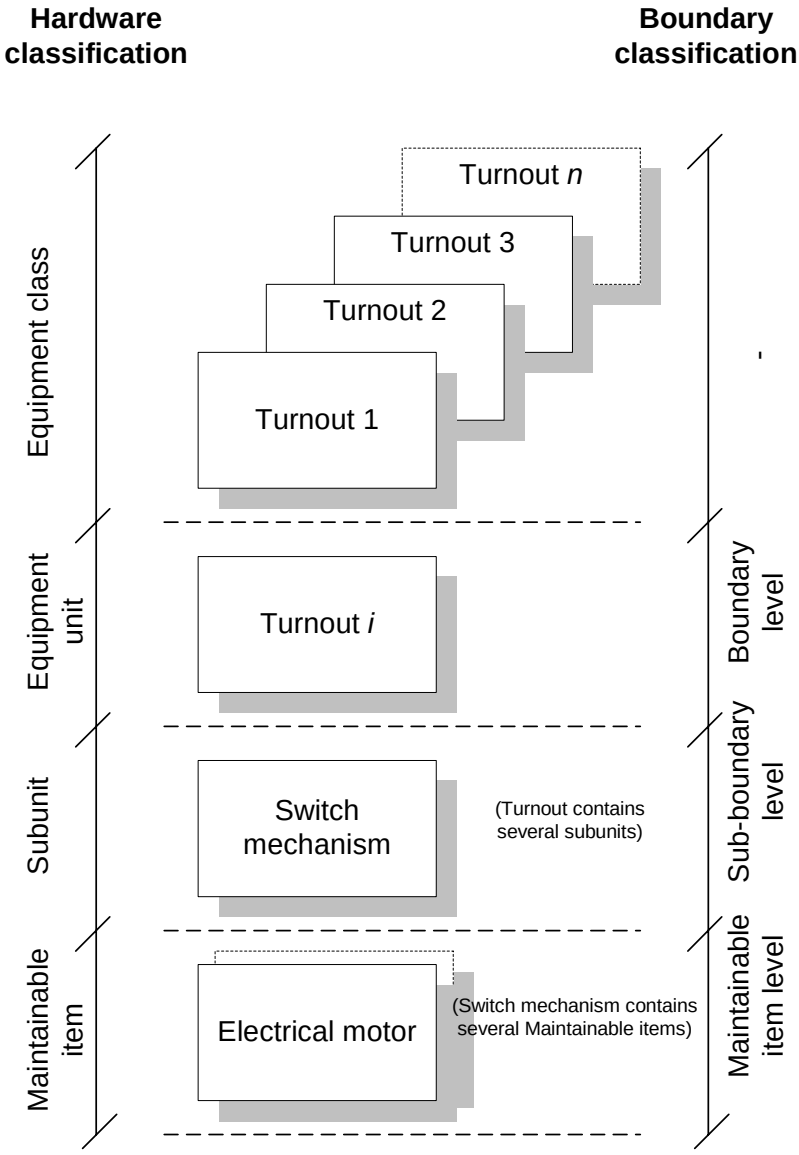


Figure 52 Example of equipment hierarchy (adapted from ISO 14224)

13.4 RAMS database structure

13.4.1 Data categories

The RAMS data shall be collected in an organised and structured way. The major data categories for equipment, failure, maintenance and state information data are given below. Note that the OREDA concept (ISO 14224) does not include state information data. In Figure 53 the inclusion of state information is explicitly demonstrated.

13.4.2 Equipment data

The description of equipment is characterised by:

1. identification data; e.g. equipment location, classification , installation data, equipment unit data;
2. design data; e.g. manufacturer's data, design characteristics;
3. application data; e.g. operation, environment.

These data categories shall in part be general for all equipment classes e.g. type classification and specific for each equipment unit e.g. radius for a turnout. This shall be reflected in the database structure. For more details see Table 11.

13.4.3 Failure data

These data are characterised by:

1. identification data, failure record and equipment location;
2. failure data for characterising a failure, e.g. failure date, maintainable items failed, severity class, failure mode, failure cause, method of observation.
3.

For more details see Table 12.

13.4.4 Maintenance data

These data are characterised by:

1. identification data; e.g. maintenance record, equipment location, failure record;
2. maintenance data; parameters characterising a maintenance, e.g. date of maintenance, maintenance category, maintenance activity, items maintained, maintenance man hours per discipline, active maintenance time, down time.

For more details see Table 14.

The type of failure and maintenance data shall normally be common for all equipment classes with exceptions where specific data types need to be collected.

Corrective maintenance events shall be recorded in order to describe the corrective action following a failure. Preventive maintenance records are required to get the complete lifetime history of an equipment unit.

13.4.5 State information

State information (condition monitoring information) may be collected in the following manners:

- Readings and measurements during maintenance
- Observations during normal operation
- Continuous measurements by use of sensor technology

13.5 Data format

Each record e.g. a failure event shall be identified in the database by a number of attributes. Each attribute describes one piece of information, e.g. the failure mode. It is recommended that each piece of information is coded where possible. The advantages of this approach versus free text are:

- queries and analysis of data are facilitated;
- ease of data input;
- consistency check undertaken at the input; by having pre-defined codes.

The range of pre-defined codes should be optimised. A short range of codes may be too general to be useful. A long range of codes may give a more precise description, but will slow the input process and may not be used fully by the data acquirer.

The disadvantage of a pre-defined list of codes versus free text is that some detailed information may be lost. It is recommended that free text is included to contain supplementary information. A free text field with additional information is also useful for quality control of data.

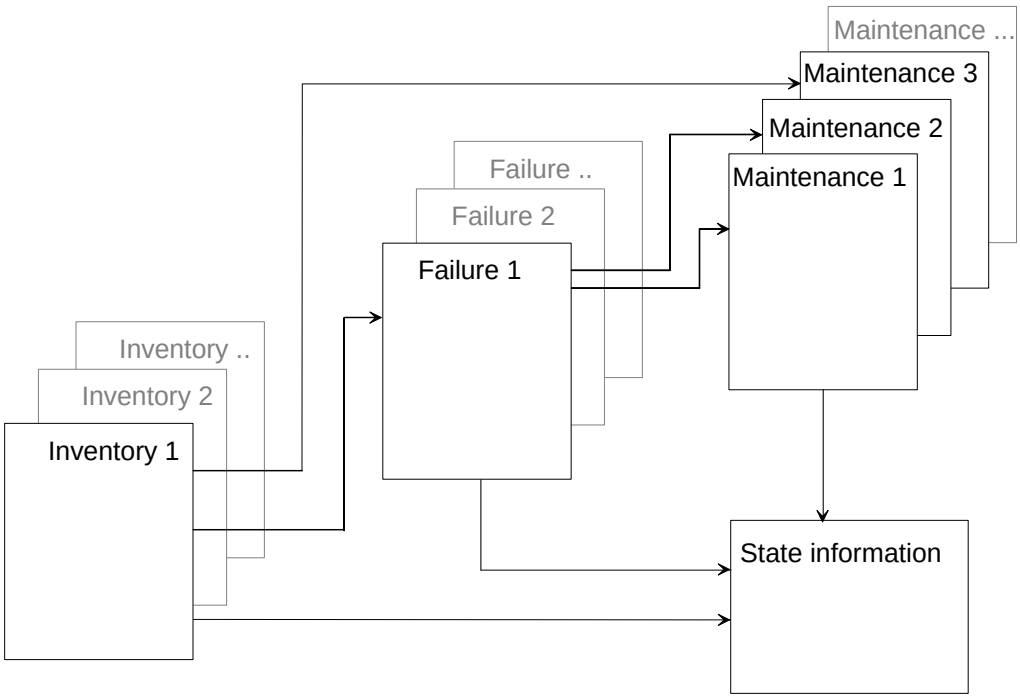


Figure 53 Logical RAMS database structure

13.6 Database structure

The data collected shall be organised and linked in a database to provide easy access for updates, queries and analysis, e.g. statistics, lifetime analysis. An example on how the information in the database may be logically given is shown in Figure 53.

13.7 Equipment, failure maintenance and state information data

13.7.1 Equipment data

The classification of equipment into technical, operational and environmental parameters is the basis for the collection of RAMS data. This information is also necessary to determine if the data is suitable or valid for various applications. There is some data which is common to all equipment classes and some data which are specific for each equipment class.

Table 11 Equipment data (Adapted from ISO 14224)

Main categories	Sub-categories	Data
Identification	Equipment location	- Equipment tag number (*)
	Classification	- Equipment unit class e.g. (*) - Equipment type (see Annex A) (*) - - Application (see Annex A)(*)
	Installation data	Country Line (from A to B) Type of line e.g. double track, high speed line Type of track e.g. main track
	Equipment unit data	- Equipment unit description (nomenclature) - Unique number e.g. serial number - Subunit redundancy e.g. no of redundant subunits
Design	Manufacturer's data	- Manufacturer's name (*) - Manufacturer's model designation (*)
	Design characteristics	- Relevant for each equipment class e.g. turnout radius, current feeder voltage, see Annex A (*)
	Cost data	
Application	Operation (normal use)	- Mode while in the operating state, e.g. continuous running, standby, normally closed/open, intermittent - Date the equipment unit was installed or date of production start-up - Surveillance period (calendar time)(*) - The accumulated operating time during the surveillance period - Number of demands during the surveillance period as applicable - Operating parameters as relevant for each equipment class e.g. number of trains passing per hour, see Annex A
	Environmental factors	External environment (severe, moderate, benign)[a]
Remarks	Additional information	- Additional information in free text as applicable
[a] Features to be considered, e.g. degree of protective enclosure, vibration, salt spray or other corrosive external fluids, dust, heat, humidity, snow.		

The minimum data needed to meet the objectives of ISO 14224 is identified by (*).

To ensure that the objectives of this International standard are met, there is a minimum of data to be collected. These data is identified by an asterisk (*) in Table 11 - Table 14.

Table 11 contains the data common to all equipment classes. In addition some data which is specific for each equipment class should be reported. Annex A gives examples of such data for some equipment classes. In the examples in Annex A high priority data is indicated.

13.7.2Failure data

A uniform definition of failure and method of classifying failures is essential when data from different sources (plants and operators) should be combined in a common RAMS database.

A common report for all equipment classes shall be used for reporting failure data. The data is given in Table 12.

Table 12 Failure data (From ISO 14224)

Category	Data	Description
Identification	Failure record (*)	Unique failure identification
	Equipment location (*)	Tag number
Failure data	Failure date (*)	Date the failure was detected (year/month/day)
	Failure mode (*)	At equipment unit level as well as at maintainable item level)
	Impact of failure on operation	See Table 13 below.
	Severity class (*)	Effect on equipment unit function: critical failure, non-critical failure
	Failure descriptor	The descriptor of the failure (see Table 19)
	Failure cause	The cause of the failure (see Table 20)
	Subunit failed	Name of subunit that failed (see examples in Annex A)
	Maintainable Item(s) failed	Specify the failed maintainable item(s) (see examples in Annex A)
	Method of observation	How the failure was detected (see Table 21)
Remarks	Additional information	Give more details, if available, on the circumstances leading to the failure, additional information on failure cause etc.

The minimum data needed to meet the objectives of the ISO 14224 is identified by (*).

Table 13 Impact of failure on operation

Description	Unit, code list or comment
Number of trains delayed less than 5 minutes	Number
Number of trains delayed between 5 and 30 minutes	Number
Number of trains delayed more than 30 minutes	Number
Period of total unavailability	Minutes
Period of reduced performance	Minutes
Safety impact?	If “Yes”, specify
Material damage?	If “Yes”, specify
Environmental impact?	If “Yes”, specify

13.7.3Maintenance data

Maintenance is carried out:

1. To correct a failure (corrective maintenance);
2. As a planned and normally periodic action to prevent failure from occurring (preventive maintenance).

A common report for all equipment classes shall be used for reporting maintenance data. The data is given in Table 14.

Table 14 Maintenance data (From ISO 14224)

Category	Data	Description
Identification	Maintenance record (*)	Unique maintenance identification
	Equipment location (*)	Tag number
	Failure record (*)	Corresponding failure identification (corrective maintenance only)
Maintenance data	Date of maintenance (*)	Date when maintenance action was undertaken
	Maintenance category	Corrective maintenance or preventive maintenance
	Maintenance activity	Description of maintenance activity (see Table 22)
	Impact of maintenance on operation	Zero, partial or total, (safety consequences may also be included)
	Subunit maintained	Name of subunit maintained (see Annex A) NOTE - For corrective maintenance, the subunit maintained will normally be identical with the one specified on the failure event report
	Maintainable item(s) maintained	Specify the maintainable item(s) that were maintained (see Annex A)
	Spare parts	Spare parts required to restore the item Cost of spare parts, or links to a cost structure database..
Maintenance resources[a]	Maintenance man-hours, per discipline	Maintenance man-hours per discipline (mechanical, electrical, instrument, others)
	Maintenance man-hours, total	Total maintenance man-hours.
Maintenance time	Active maintenance time	Time duration for active maintenance work on the equipment
	Down time	The time interval during which an item is in a down state
Remarks	Additional information	Give more details, if available, on the maintenance action, e.g. abnormal waiting time, relation to other maintenance tasks

13.7.4State information

State information (condition monitoring information) may be collected in the following manners:

- Readings and measurements during maintenance
- Observations during normal operation
- Continuous measurements by use of sensor technology

Table 15 State information, discrete readings

Category	Data	Description
Identification	State information record	Unique state information identification
	Equipment location	Tag number
	Maintenance record	Corresponding maintenance identification, i.e. an observation is recorded either related to corrective or preventive maintenance
	Failure record	Corresponding failure identification (if no maintenance is performed in relation to the failure)
State information	Date of observation	Date when state information was read
	Type of measurement	What measurement is obtained? For example a distance measure,
	Value	What are the readings of the measurement?
Remarks	Additional information	Give more details
If the readings are taken during normal operation, there will not be a corresponding maintenance or failure record. In this case the state information is linked directly to the inventory record		

Table 16 State information, continuous readings

Category	Data	Description
Identification	State information record	Unique state information identification
	Equipment location	Tag number
	Type of measurement	What measurement is obtained? For example a distance measure,
State information	Sampling frequency	What is the sampling frequency?
	Sensor	What type of sensor is used
	Data compression principle	How is data compressed, e.g. Fast Fourier Transform
Remarks	Additional information	Give more details
State information is linked directly to the inventory record for continuous readings		

Table 17 Example of breakdown into maintainable items (turnouts)

Equipment unit	Turnout				
Subunit	Switch mechanism	Rails	Sleepers	Interface/-fastening	Miscel-laneous
Maintainable items	Motor Moving rods Switch locks Detector rod	Stock rail Switch rail Check rail Crossing point	Concrete sleepers Wooden sleepers	Heel blocks Distance blocks Slide plates Sole plate Chair/base-plates fastening Spring clip	

Table 18 Example failure modes at maintainable item level (turnouts)

Item	Code	Definition	Description
Turnout	FTO	Fail to open	Fail to move to a “turnout position” (from a straight position)
	FTC	Fail to close	Fail to move (back) to a straight position
	SOP	Spurious opening	Moves to a “turnout position” without any demand
	SCL	Spurious closure	Moves to a straight position without any demand
	IMP	Intermediate position	The switch is in a position between open and closed
	USP	Unsafe passage	The turnout cannot be passed in a safe manner, e.g. check rails out of position.
Motor	NOE	No effect	No effect from the motor
	REE	Reduced effect	Reduced performance of the motor
Moving rods	STU	Stuck	
Switch locks			
..			
..			

13.8 Failure and maintenance notations

In this chapter proposed code lists for the following topics are provided:

- Failure descriptor (Physical failure cause)
- Failure cause (Root causes related to design, specification, organisation etc.)
- Method of detection
- Maintenance activity

Note the lists are considered to be general and are common to all equipment classes relevant for railway applications.

Table 19 Failure descriptors (From ISO 14224)

No.	Notation	Description
1.0	Mechanical failure-general	A failure related to some mechanical defect, but where no further details are known
1.1	Leakage	External and internal leakages, either liquids or gases. If the failure mode at equipment unit level is leakage, a more causal oriented failure descriptor should be used wherever possible
1.2	Vibration	Abnormal vibration. If the failure mode at equipment level is vibration, a more causal oriented failure descriptor should be used wherever possible
1.3	Clearance/ alignment failure	Failure caused by faulty clearance or alignment
1.4	Deformation	Distortion, bending, buckling, denting, yielding, shrinking, etc.
1.5	Looseness	Disconnection, loose items
1.6	Sticking	Sticking, seizure, jamming due to reasons other than deformation or clearance/alignment failures
2.0	Material failure-general	A failure related to a material defect, but no further details known
2.1	Cavitation	Relevant for equipment such as pumps and valves
2.2	Corrosion	All types of corrosion, both wet (electrochemical) and dry (chemical)
2.3	Erosion	Erosive wear
2.4	Wear	Abrasive and adhesive wear, e.g. scoring, galling, scuffing, fretting, etc.
2.5	Breakage	Fracture, breach, crack
2.6	Fatigue	If the cause of breakage can be traced to fatigue, this code should be used
2.7	Overheating	Material damage due to overheating/burning
2.8	Burst	Item burst, blown, exploded, imploded, etc.
3.0	Instrument failure – general	Failure related to instrumentation, but no details known
3.1	Control failure	
3.2	No signal/indication/-alarm	No signal/indication/alarm when expected
3.3	Faulty signal/indication/-alarm	Signal/indication/alarm is wrong in relation to actual process. Could be spurious, intermittent, oscillating, arbitrary
3.4	Out of adjustment	Calibration error, parameter drift
3.5	Software failure	Faulty or no control/monitoring/operation due to software failure
3.6	Common mode failure	Several instrument items failed simultaneously, e.g. redundant fire and gas detectors
4.0	Electrical failure-general	Failures related to the supply and transmission of electrical power, but where no further details are known
4.1	Short circuiting	Short circuit
4.2	Open circuit	Disconnection, interruption, broken wire/cable
4.3	No power/ voltage	Missing or insufficient electrical power supply
4.4	Faulty power/voltage	Faulty electrical power supply, e.g. over voltage
4.5	Earth/isolation fault	Earth fault, low electrical resistance
5.0	External influence – general	The failure where caused by some external events or substances outside boundary, but no further details are known
5.1	Blockage/plugged	Flow restricted/blocked due to fouling, contamination, icing, etc.

No.	Notation	Description
5.2	Contamination	Contaminated fluid/gas/surface e.g. lubrication oil contaminated, gas detector head contaminated
5.3	Miscellaneous external influences	Foreign objects, impacts, environmental, influence from neighbouring systems
6.0	Miscellaneous – general[a]	Descriptors that do not fall into one of the categories listed above.
6.1	Unknown	No information available related to the failure descriptor.

[a] The data acquirer shall judge which is the most important descriptor if more than one exist, and try to avoid the 6.0 and 6.1 codes.

Table 20 Failure causes (From ISO 14224)

No.	Notation	Description
1.0	Design related causes - general	Failure related to inadequate design for operation and/or maintenance, but no further details known
1.1	Improper capacity	Inadequate dimension/capacity
1.2	Improper material	Improper material selection
1.3	Improper design	Inadequate equipment design or configuration (shape, size, technology, configuration, operability, maintainability, etc.)
2.0	Fabrication/installation related causes - general	Failure related to fabrication or installation, but no further details known
2.1	Fabrication error	Manufacturing or processing failure
2.2	Installation error	Installation or assembly failure (assembly after maintenance not included)
3.0	Failures related to operation/maintenance - general	Failure related to the operation/use or maintenance of the equipment, but no further details known
3.1	Off-design service	Off-design or unintended service conditions e.g. compressor operation outside envelope, pressure above specification, etc.
3.2	Operating error	Mistake, misuse, negligence, oversights, etc. during operation
3.3	Maintenance error	Mistake, errors, negligence, oversights, etc. during maintenance
3.4	Expected wear and tear	Failure caused by wear and tear resulting from normal operation of the equipment unit
4.0	Failures related to administration - general	Failure related to some administrative system, but no further details known
4.1	Documentation error	Failure related to procedures, specifications, drawings, reporting, etc.
4.2	Management error	Failure related to planning, organisation, quality control/assurance, etc.
5.0	Miscellaneous - general [a]	Causes that do not fall into one of the categories listed above.
5.1	Unknown [a]	No information available related to the failure cause.

[a] The data acquirer shall judge which is the most important cause if more than one exist, and try to avoid the 5.0 and 5.1 codes.

Table 21 Method of detection (From ISO 14224)

No.	Notation	Description
1	Preventive maintenance	Failure discovered during preventive service, replacement or overhaul of an item when executing the preventive maintenance program.
2	Functional testing	Failure discovered by activating an intended function and comparing the response against a predefined standard.
3	Inspection	Failure discovered during planned inspection e.g. visual inspection, non-destructive testing
4	Periodic condition monitoring	Failures revealed during a planned, scheduled condition monitoring of a predefined failure mode, either manually or automatically e.g. thermography, vibration measuring, oil analysis, sampling
5	Continuous condition monitoring	Failures revealed during a continuous condition monitoring of a predefined failure mode.
6	Corrective maintenance	Failure observed during corrective maintenance
7	Observation	Observation during routine or casual non-routine operator checks mainly by senses (noise, smell, smoke, leakage, appearance, local indicators)
8	Combination	Several of above methods involved. If one of the methods is the predominant one, this should be coded.
9	Production interference	Failure discovered by production upset, reduction, etc.
10	Other	Other observation method

Table 22 Maintenance activity (From ISO 14224)

No.	Activity	Description	Examples	Use[a]
1	Replace	Replacement of the item by a new, or refurbished, of the same type and make	Replacement of a worn-out bearing	C, P
2	Repair	Manual maintenance action performed to restore an item to its original appearance or state	Repack, weld, plug, reconnect, remake, etc.	C
3	Modify	Replace, renew, or change the item, or a part of it, with an item/part of different type, make, material or design	Install a filter with smaller mesh diameter, replace a lubrication oil pump with another type etc.	C
4	Adjust	Bringing any out-of-tolerance condition into tolerance	Align, set and reset, calibrate, balance	C
5	Refit	Minor repair/servicing activity to bring back an item to an acceptable appearance, internal and external	Polish, clean, grind, paint, coat, lube, oil change, etc.	C
6	Check[b]	The cause of the failure is investigated, but no maintenance action performed, or action deferred. Able to regain function by simple actions, e.g. restart or resetting	Restart, resetting, etc. In particular relevant for functional failures e.g. fire and gas detectors	C
7	Service	Periodic service tasks. Normally no dismantling of the item	E.g. cleaning, replenishment of consumables, adjustments and calibrations	P
8	Test	Periodic test of function availability	Function test of fire pump, gas detector etc.	P
9	Inspection	Periodic inspection/check. A careful scrutiny of an item carried out with or without dismantling, normally by use of senses	All types of general *checks*. Includes minor servicing as part of the inspection task	P
10	Overhaul	Major overhaul	Comprehensive inspection/overhaul with extensive disassembly and replacement of items as specified or required	P(C)
11	Combination	Several of the above activities are included	If one activity is the dominating, this could alternatively be recorded	C, P
12	Other	Other maintenance activity than specified above		C, P

[a] C = used typically in corrective maintenance, P = used typically in preventive maintenance.

[b] "Check" includes both circumstances where a failure cause was revealed, but no maintenance action considered necessary, and where no failure cause could be found.

14. COLLECTION AND ANALYSIS OF RELIABILITY DATA

Collection and analysis of reliability data is an important element of maintenance management and continuous improvement. There are several aspects of utilizing experience data and we will in the following focus on:

- Learning from experience. That is, when a problem occurs the failure and maintenance databases can be searched for events which are similar to the current problem. If the database is properly updated, we might then find information about solutions that proved to be efficient, and also solutions that did not proved to be efficient in the past.
- Identification of common problems. By producing "Top ten"-lists (visualised by Pareto diagrams) the database can be used to identify common problems. For example which component contribute most to the total downtime (cost drivers), what are the dominate failure causes etc. "Top-ten" lists are used as a basis for deciding where to spend resources for improvements.
- A basis for estimation of reliability parameters. Important parameters to use in RAMS analyses are the Mean Time To Failure (MTTF), ageing parameters, P-F intervals and repair times.

14.1 Short introduction to various types of analyses

14.1.1 Learning from experience

The database may be used as a "case based" experience database, i.e. each failure and maintenance report represents a case from which experience might be gained. To utilise the information it is important that the failure and maintenance reports contain extensive information about the failure, the causes of the failures, what corrective actions were made, and also the results of any corrective action taken.

Since the database contains thousands of records it is also important that it is easy to search the database for relevant cases. The use of pre-defined lists in the database will make such search easier. In addition to such features built into the database, it is also important that the database can easily be searched. Most database systems have "search engines" for identification of relevant records. The search criteria can either be specified by a user friendly dialogue, or by some command statement such as an SQL statement.

In a practical situation when a problem is at hand, one will typical search for "similar" problems. It is however, not a straight forward task to define "similar" in this context. A problem is often characterised by a set of "attributes". However these attributes are on different levels of measurements (see Section 14.2.1 page 129) and the definition of "similarity" measures is therefore complicated. Several techniques for identification of similar events are described under the broad class of "data mining" techniques, see e.g. Fayyad *et al*. (1996). Data mining is further one part of the more general problem of Knowledge Discovery in Databases (KDD) defined as: "the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad *et al*. 1996, p 6).

Identification of common problems

A database is also a useful source for identification of common problems. The idea is to identify those problems which contribute most to the threat against safety, punctuality/availability, costs etc. This process is often carried out in two or more steps. First

the database is searched for components contributing much to for example delay time. Thereafter these components/systems are further investigated to identify failure causes. A so-called Pareto diagram is often produced to visualise the result of the "Top ten list". An example is shown in Figure 54.

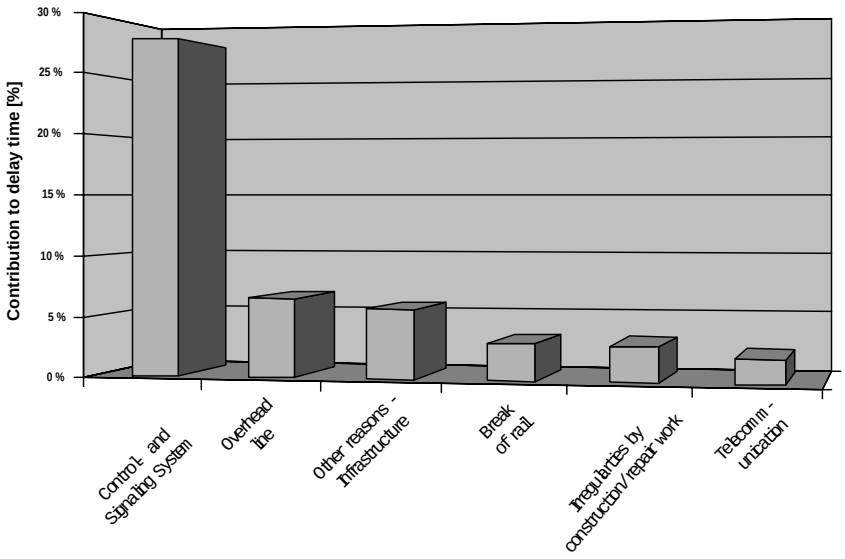


Figure 54 Pareto diagram showing contribution to delay time

Very often the two or three first "bars" account for a large amount of variable of interest. When constructing the Pareto diagram the following dimensions should be considered:

- What should be the "score"-variable?
- What is the "grouping" variable?

The "score" variable

The "score" variable represents the cost in some way or another. Various information from the failure and maintenance database can be used to produce a "score" variable, e.g.:

- Severity class
- Impact on failure on operation (number of trains affected, safety impact, material damages etc)
- Downtime
- Spare part consumption (costs)
- Maintenance man-hours

One or more of these variables should be combined into one quantitative measure representing the "score" for each event in the failure/maintenance database. This "score" variable is used when producing the Pareto diagram.

The "grouping" variable

The equipment class is usually the first variable to group on. Now several paths of breakdown exist. For example a breakdown into equipment types and/or application may be performed. Another breakdown is to group on sub-units and/or maintainable items.

14.1.2A basis for estimation of reliability parameters

Reliability parameters are important input maintenance optimisation methodology the following parameters are of most importance:

- Parameters for “non-observable” failure progression
  - Mean time to failure MTTF (inverse of the failure rate)
  - Ageing parameter (α)
- Parameters for “observable” failure progression
  - P-F intervals
  - Parameters describing the “failure limit”
- Other parameters
  - Mean time to repair
  - Spare part consumption
  - Mean down time when a failure has occurred

In the first situation we will take advantages of standard life time analysis which will be covered in section 14.5.

14.2 Simple plotting techniques

In this chapter we present some basic methods for playing around with the data. The techniques may be used to get a good overview over the data, identify important explanatory variables etc. These methods are found in most commercial statistical packages. First we give a definition of different levels of measurements.

14.2.1Levels of measurements

Data can be measured on several levels. The traditional classification of levels of measurements was developed by Stevens (1946). He identified four levels; nominal, ordinal, interval and ratio.

Nominal-Level Measurement

The “lowest” level in Stevens’ typology is the nominal level. No ordering between the values of the variable is assumed. This level is typically used for categorical data.

Ordinal-Level Measurement

The ordinal level is used when it is possible to rank-order all categories according to some criterion. Note that the ordinal level only rank-orders the values. It is not possible to say anything about *how much* the difference between low, medium and high is.

Interval-Level Measurement

In the interval level situation there is an ordering of the categories, in addition the distance between the categories are defined in terms of fixed and equal units. The temperature

(measured in °C or °F) is a typical example. For the interval level there is no *fixed* zero point. Thus it does not make sense to claim that 20 °C is twice as hot as 10 °C.

Ratio-Level Measurement

The "highest" level in Stevens' typology is the ratio level. The ratio level has the same properties as the interval level. In addition there is defined a fixed zero point. For example temperature measured in °K satfy the properties of a ratio level measurement. Also pressure measured in Bar will satfy the ratio level measurement.

When analysing data it is important to be aware of the level at which the data is measured. Parametric methods are usually based on data measured on the interval or ratio level.

14.2.2Bar charts

A Bar chart displays a bar for each category of a variable. Generally, bar charts display counts of each category of a qualitative variable (either numeric or character) or means of a quantitative variable grouped by a qualitative variable.

14.2.3Pie charts

A Pie chart displays a pie divided into pieces. Each piece corresponds to a category of a variable.

14.2.4Box-and-whiskers plots

A Box-and-whiskers plot shows the distribution of a quantitative variable. For a plot of a quantitative variable grouped by a qualitative variable, the distribution within each category may be displayed to show differences between groups. An example of a box and whiskers plot is shown in Figure 55.

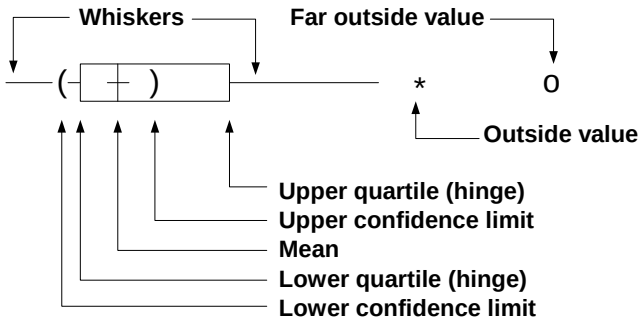


Figure 55 Example of box and whiskers plot

The vertical line inside the box represents the median and the vertical ends of the box represent the lower and upper hinges (the 25^{th} and 75^{th} percentiles). In addition, the following represent:

Asterisks	Outside values, which are data values outside the inner fences. Where Hspread is the absolute value of the difference between the two hinges, inner fences are defined as: Lower fence = lower hinge - 1.5(Hspread) Upper fence = upper hinge + 1.5(Hspread)
Open circles	Far outside values, which are data values outside the outer fences. Outer fences are defined as: Lower fence = lower hinge - 3.0(Hspread) Upper fence = upper hinge + 3.0(Hspread)

14.3 Qualitative analysis

14.3.1 Total maintenance cost

In order to control maintenance cost it is important to identify the "cost drivers". The cost may usually be measured in terms of one or more of the following variable

- Severity class (in failure database)
- Impact on failure on operation (number of trains affected, safety impact, material damages etc)
- Downtime
- Spare part consumption (costs)
- Maintenance man-hours

The Pareto diagram may be used to show the relative contribution from various components to one of the cost variables listed above. When analysing the data it is important to understand the database structure. The "score" variable will typically be in either the failure or maintenance databases, whereas the grouping variable (equipment class) is defined in the inventory database.

14.3.2 Failure cause analysis

The main objective of the failure cause analysis is to identify failure causes that repeat themselves. The recommended procedure is to start with equipment identified in the "upper ten" lists of Section 14.3.1. For those equipment classes that contribute much to the total maintenance cost, the most important failure causes are identified also by means of "Pareto diagrams". Note that failure causes often are specified at two levels[7]:

- Failure descriptor (Physical failure cause)
- Failure cause (Root causes related to design, specification, organisation etc.)

The physical failure cause will often be the starting point in a maintenance analysis, since the main objective of the maintenance *tasks* is to prevent these failure causes from leading to a failure. However, in many situation the most efficient approach is to start with the root causes, since they by definition are the primarily source of the problem.

When a specific failure cause has been identified it is often convenient to list corresponding failure and maintenance reports to get a better understanding of what the problem really is.

[7] E.g. in OREDA, see ISO 14224 (1999).

The narrative information in the maintenance report may often be very valuable in order find solutions to frequent problem.

14.4 Estimation procedures for a constant failure rate

14.4.1 Objective

The objective of this section is to describe methods for obtaining failure rate estimates in situation where the failure rate is assumed to be constant, i.e. no ageing effects. Even if this assumption does not hold, it might be valuable to have this as a starting point in order to get an overview of the reliability characteristics of various components. Note that in the situation of constant failure rate, there is an inverse relationship between the failure rate and the mean time to failure. In Section 14.5 we will discuss more advanced methods that might be used in case of a non-constant failure rate.

14.4.2 Estimators and Uncertainty Limits for a Homogeneous Sample

When we have failure data from identical items that have been operating under the same operational and environmental conditions, we have a so-called *homogeneous sample*. The only data we need to estimate the failure rate λ in this case, are the observed number of failures, n, and the aggregated time in service, t.

The estimator of λ is given by:

$$\hat{\lambda} = \frac{\text{Number of failures}}{\text{Aggregated time in service}} = \frac{n}{t} \tag{97}$$

See e.g. Rausand and Høyland (2003) for further details.

Note that this approach is valid only in the following situations:

- Failure times for a specified number of items, with the *same* failure rate λ, are available.
- Data (several failures) is available for *one* item for a period of time, and the failure rate λ is constant during this period.
- A combination of the two above situations, i.e., there are several items where each item might have several failures. This is the typical situation for most reliability databases.

Similarly, if we want an estimate for MTTF, we may set

$$\hat{MTTF} = \frac{\text{Aggregated time in service}}{\text{Number of failures}} = \frac{t}{n} \tag{98}$$

Uncertainty intervals for the failure rate

The uncertainty of the failure rate estimate may be presented as a 90% confidence interval. This is an interval (λ_L,λ_U), such that the "true value" of λ fulfils:

$$\Pr(\lambda_L \le \lambda < \lambda_U) = 90\% \tag{99}$$

With n failures during an aggregated time in service t, this 90% confidence interval is given by:

$$\left(\frac{1}{2t} z_{0.95,2n}, \frac{1}{2t} z_{0.05,2(n+1)} \right) \tag{100}$$

where $z_{0.95,\nu}$ and $z_{0.05,\nu}$ denote the *upper* 95% and 5% percentiles, respectively, of the χ^2-distribution with ν degrees of freedom, see Table 27, page 148.

Example 14.1

Assume that n = 6 failures have been observed during an aggregated time in service t = 10 000 hours.

The failure rate estimate is then given by:

$\hat{\lambda} = n/t = 6 \cdot 10^{-4}$ failures per hour

and a 90% confidence interval is given by:

$$\left(\frac{1}{2t} z_{0.95,2n}, \frac{1}{2t} z_{0.05,2(n+1)} \right) = \left(\frac{1}{20000} z_{0.95,12}, \frac{1}{20000} z_{0.05,14} \right) = (2.6 \cdot 10^{-4}, 11.8 \cdot 10^{-4})$$

The estimate and the confidence interval are illustrated in Figure 56.

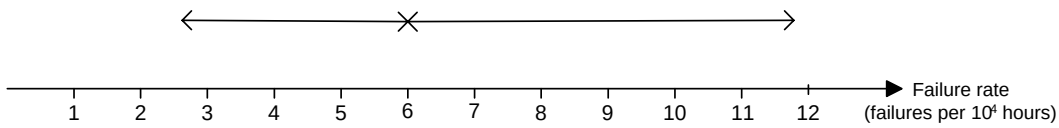


Figure 56 Estimate and 90% Confidence Interval

□

Note
The given interval is a confidence interval for the failure rate for the items we have data for. There is no guarantee that items installed in the future will have a failure rate within this interval.

14.4.3Multi-Sample Problems

In many cases we do not have a homogeneous sample of data. The aggregated data for an item may come from different installations with different operational and environmental conditions, or we may wish to present an "average" failure rate estimate for slightly different items. In these situations we may decide to merge several more or less homogeneous samples, into what we call a multi-sample.

The various samples may have different failure rates, and different amounts of data - and thereby different confidence intervals. This is illustrated in Figure 57.

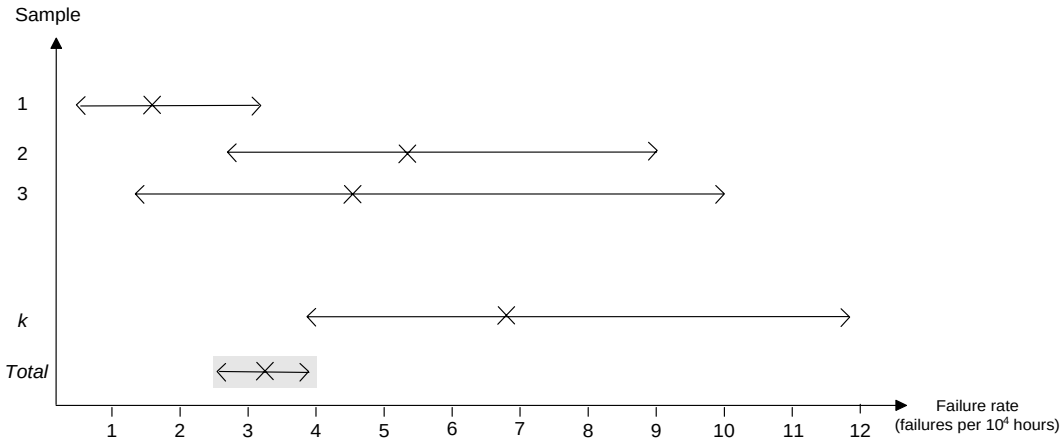


Figure 57 Multi-Sample Problem

To merge all the samples and then estimate the "average" failure rate as the total number of failures divided by the aggregated time in service will not always give an adequate result. The "confidence" interval will especially be unrealistically short, as illustrated in Figure 6. We therefore need a more advanced estimation procedure to take care of the multi-sample problem.

Below, the so-called OREDA-estimator of the "average" failure rate in a multi-sample situation is presented together with a 90% uncertainty interval. Spjøtvoll (1985) gives a rationale for the estimation procedure.

The OREDA-estimator is based on the following assumptions:

- We have k different samples. A sample may e.g., correspond to a platform, and we may have data from similar items used on k different platforms.
- In sample no. i we have observed n_i failures during a total time in service t_i, for i =1,2,…, k.
- Sample no. i has a constant failure rate λ_i, for i =1,2,…, k.
- Due to different operational and environmental conditions, the failure rate λ_i may vary between the samples.

The variation of the failure rate between samples may be modelled by assuming that the failure rate is a random variable with some distribution given by a probability density function $\pi(\lambda)$.

The mean, or "average" failure rate is then: $\theta = \int_0^\infty \lambda \cdot \pi(\lambda)\,\mathrm{d}\lambda$.

and the variance is: $\sigma^2 = \int_0^\infty (\lambda - \theta_\Lambda)^2 \cdot \pi(\lambda)\,\mathrm{d}\lambda$.

To calculate the multi-sample OREDA-estimator, the following procedure is used:

1. Calculate an initial estimate for the mean ("average") failure rate θ, by pooling the data:

$$\hat{\theta}_1 = \frac{\text{Total no. of failures}}{\text{Total time in service}} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i}$$

2. Calculate:

$$S_1 = \sum_{i=1}^{k} t_i$$

$$S_2 = \sum_{i=1}^{k} t_i^2$$

$$V = \sum_{i=1}^{k} \frac{(n_i - \hat{\theta}_1 t_i)^2}{t_i} = \sum_{i=1}^{k} \frac{n_i^2}{t_i} - \hat{\theta}_1^2 S_1$$

3. Calculate an estimate for σ^2, a measure of the variation between samples, by:

$$\hat{\sigma}^2 = \frac{V - (k-1)\hat{\theta}_1}{S_1^2 - S_2} \times S_1$$

If $\hat{\sigma}^2 \leq 0$, we estimate the variation between samples by $\hat{\sigma}^2 = \frac{1}{k-1} \sum_{i=1}^{k} \left(\frac{n_i}{t_i} - \hat{\theta}_1 \right)^2$.

4. Calculate the final estimate θ^* of the mean ("average") failure rate θ by:

$$\theta^* = \frac{1}{\sum_{i=1}^{k} \frac{1}{\frac{\hat{\theta}_1}{t_i} + \hat{\sigma}^2}} \times \sum_{i=1}^{k} \left(\frac{1}{\frac{\hat{\theta}_1}{t_i} + \hat{\sigma}^2} \times \frac{n_i}{t_i} \right)$$

5. Let $SD = \hat{\sigma}$
6. The *lower* and *upper* "uncertainty" values are given by

$$\int_{Lower}^{Upper} \pi(\lambda)\, d\lambda = 90\%$$

Since the distribution $\pi(\lambda)$ is not known in advance, the following pragmatic approach is used:

7. $\pi(\lambda)$ is assumed to be the probability density function of a *gamma* distribution with parameters α and β.

8. The parameters α and β are estimated by:

$$\hat{\beta} = \frac{\theta^*}{\hat{\sigma}^2}$$
$$\hat{\alpha} = \hat{\beta} \cdot \theta^*$$

9. The following formulas are now applied:

$$Lower = \frac{1}{2\hat{\beta}} z_{0.95,2\hat{\alpha}}$$
$$Upper = \frac{1}{2\hat{\beta}} z_{0.05,2\hat{\alpha}}$$

where $z_{0.95,\nu}$ and $z_{0.05,\nu}$ denote the *upper* 95% and 5% percentiles, respectively, of the χ^2-distribution with ν degrees of freedom, see Table 27, page 148. In situations where ν is not an integer, an interpolation in the χ^2-distribution is performed.

14.5 Life time data analysis

14.5.1 Objective

The primary objective of life data analysis is to obtain information about the life distribution $F(t)$ for a unit. The lifetime of a unit is defined as the time from the unit is installed until it fails, i.e. it is not able to perform the intended function. Before a unit is installed the lifetime T is not known in advanced, but treated as a random variable with distribution function $F(t) = P(T \leq t)$. In addition to the distribution function $F(t)$, the failure rate function $z(t)$ is of great interest.

In maintenance optimisation we are especially focusing on the ageing parameter which are of crucial importance when determining the optimum maintenance interval. The form of the failure rate function will indicate whether there are strong ageing or not.

If lifetimes of several units are available it is possible to fit parametric life distributions to the failure data. Examples of such lifetime distributions are:

- The exponential distribution
- The Weibull distribution
- The gamma distribution
- The lognormal distribution
- The inverse Gaussian distribution

The parametric forms of various life distributions are described in the literature, see e.g. Rausand and Høyland (2003). The estimation of parameters in these distributions requires Maximum Likelihood procedures. Now let θ be the parameter vector of interest, for example $\theta = [\alpha,\lambda]$ if the Weibull distribution is considered. Further let t_j denote the observed life times, both censored and real life times. The likelihood function is now given by:

$$L(\boldsymbol{\theta};\mathbf{t}) = \prod_{j\in C_L} F(t_j;\boldsymbol{\theta}) \prod_{j\in U} f(t_j;\boldsymbol{\theta}) \prod_{j\in C_R} R(t_j;\boldsymbol{\theta}) \quad (101)$$

where C_L, U and C_R are the set of left-censored, uncensored and right-censored life times respectively. The maximum likelihood estimate (MLE) for $\boldsymbol{\theta}$ is now the $\boldsymbol{\theta}$-vector that maximises Equation (101). Numerical methods are generally required to carry out the maximisation.

Example 14.2

Consider a situation were we have observed n failure times. The failure times are assumed to be exponentially distributed with parameter λ. The observed failure times are denoted t_1, t_2, …, t_n. Using Equation (101) the likelihood function is thus given by:

$$L(\lambda;t_1,t_2,\ldots,t_n) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i}$$

Since $L(\cdot)$ is a monotonically increasing function of the argument λ, we could maximize the logarithm of $L(\cdot)$ rather than $L(\cdot)$ which is more convenient from a mathematical point of view, i.e.

$$l(\lambda;t_1,t_2,\ldots,t_n) = \ln L(\lambda;t_1,t_2,\ldots,t_n) = n\ln\lambda - \sum_{i=1}^{n} \lambda t_i$$

By derivation wrt λ we easily obtain the maximum likelihood estimate (MLE):

$$\hat{\lambda} = n \Big/ \sum_{i=1}^{n} t_i$$ □

Exercise 22

Derive the MLE for the parameters in the Weibull distribution. Hint: it is not possible to find the solution on closed forms, i.e. an iterative procedure is required. □

14.5.2Basic model assumptions for life data analysis

Totally n units are activated in order to record their lifetimes. The units are identical, and operated under identical and independent environmental stresses. Under these conditions it is reasonable to believe that the lifetimes are independent and identically distributed (*i.i.d.*).

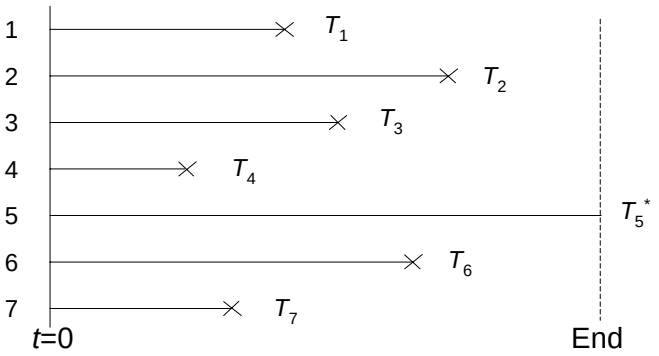


Figure 58 Conceptual model: Life data analysis

In Figure 58 the lifetimes are denoted $T_1,T_2,T_3,..,T_7$ are the lifetimes. The lifetime T_5 is a censoring lifetime, see discussion below.

$T_{(1)},T_{(2)},T_{(3)},..$ are the *ordered* lifetimes, i.e. $T_{(1)} \leq T_{(2)} \leq T_{(3)} \leq .. \leq T_{(n)}$.. For the analysis the original ordering of the lifetimes are not required and the ordered lifetimes are sufficient. This is, however, only true if the *i.i.d.* assumption holds. To check whether the *i.i.d.* assumption holds, the construction of the Nelson Aalen plot (see section 14.6.3) will be the first step.

Censoring lifetimes

A "Right" censoring lifetime means that either 1) the unit has been discarded from the experiment for some reason, or 2) the unit has not failed at the termination of the experiment.

A "Left" censoring lifetime means that it is not known when the unit has been activated, but it has been observed a period of time *T*, and then it failed. See Figure 59.

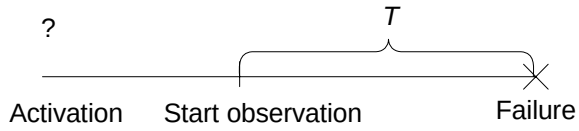


Figure 59 Example of left censoring

A "Double" censoring lifetime means that it is not known when the unit has been activated, and it is observed a period of time *T*, and it has not failed during the time of observation.

How to create *i.i.d.* data from a reliability database.

The models for life data analysis are developed for the ideal experimental situations where *n* units are put to test in order to record their lifetimes. However, this will not always be the case for data in the most reliability databases, and the models and analysis techniques must therefore be used with care since the assumptions for life data analysis may not hold. Below, some principal issues are discussed.

Example 14.3

Consider a "socket" where the unit in the socket is replaced upon a failure, and a new unit is assumed to be identical with prior units. The socket is observed in a period of time from *a* to *b*.

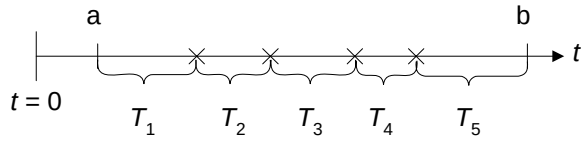


Figure 60 Lifetimes in Example 14.3

In the current framework $t = 0$ corresponds to the installation date of the unit. *a* corresponds to the surveillance start date, and *b* corresponds to the surveillance end date, i.e. [*a*,*b*] is the surveillance period.

In Figure 60, T_1 is a left censoring lifetime, T_2, T_3 and T_4 are ordinary lifetimes, and finally T_5 is a right censoring lifetime. (See also Figure 58). □

Example 14.4

Consider the following failure modes:
FTC: Fail to close
SPO: Spurious operation (closure)

We assume that the valve is replaced independent of which failure mode occurred. Other failure modes are not assumed to affect the reliability of the valve.

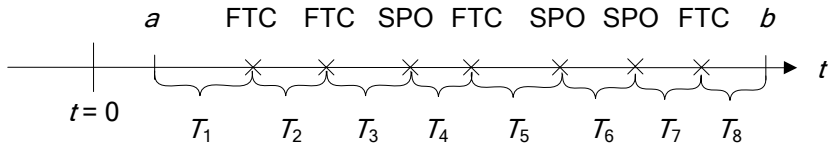


Figure 61 Lifetimes in Example 14.4

With respect to failure mode FTC we have:

Lifetimes: T_2, T_4 and T_7
Left censoring lifetime: T_1
Right censoring lifetime: T_3, T_5, T_6 and T_8

With respect to failure mode SPO we have:

Lifetimes: T_3, T_5 and T_6
Double censoring lifetime: T_1
Right censoring lifetimes: T_2, T_4, T_7 and T_8 □

Example 14.5 - Pre and post filtering

Consider the following situation and failure modes:
FMA: Failure mode of interest
FMB: Failure mode for which repair does not affect failure mode FMA, i.e. the unit is *minimal repaired,* and thus *not* repaired to an as good as new condition with respect to FMA. (In this situation a pre-filtering is appropriate).
FMC: Failure mode FMC is not a failure mode of interest, but the unit is repaired to an as good as new condition with respect to failure mode FMA. (In this situation a post-filtering is appropriate).

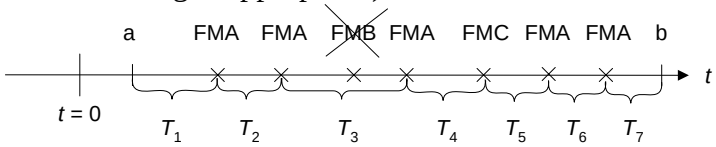


Figure 62 Lifetimes in Example 14.5

In Figure 62 $t = 0$ corresponds to the installation date of the unit, and the interval $(a,b]$ is the surveillance period. When creating an appropriate data set we first have to do some pre-filtering in order to remove the failure corresponding to failure mode FMB. This is marked with a "cross" in Figure 62. Next we have to do a post-filtering to define the failure mode

FMC as a censoring lifetime (with respect to failure mode FMA). After the pre and post filtering we have:

Lifetimes: T_2, T_3, T_5 and T_6
Left censoring lifetime: T_1
Right censoring lifetimes: T_4 and T_7 □

Using more than one "inventory" in life data analysis

In the discussion so far, the failures have been created from *one* inventory only. If it is reasons to believe that several inventories are almost similar, we can pool data from these inventories to enlarge the amount of data. The discussion above still applies, but now several inventories are used to generate the lifetimes. In order to obtain reasonable output, the *i.i.d.* assumption must still hold. That means that the units (inventories) must be similar, and operated under similar environmental conditions. A first approach to check this assumption is to perform a cross-tabulation analysis of the failure rate. Fields to consider in a cross-tabulation are:

- Taxonomy code
- Model
- Manufacturer
- Function

In the OREDA Data analysis project (Vatn 1993) more advanced methods are suggested for checking similarities between groups of inventories.

14.5.3TTT-plot

The main objective of Total Time on Test (TTT) plotting is to reveal whether the underlying failure distribution is IFR (increasing failure rate), or DFR (decreasing failure rate). It is fundamental that the *i.i.d.* assumption holds. If we have altogether n failure times, we assume:

$T_1, T_2, .., T_n \sim i.i.d.$ Further let $T_{(1)}, T_{(2)}, .., T_{(n)}$ denote the ordered lifetimes. The total time on test (TTT) at time t is defined by:

$$TTT(t) = \sum_{j=1}^{i} T_{(j)} + (n - i)t$$

where i is such that $T_{(i)} \leq t < T_{(i+1)}$

The TTT-plot is obtained by plotting

$$\left(\frac{i}{n}, \frac{TTT(T_{(i)})}{TTT(T_{(n)})} \right) \quad i = 1, .., n$$

The shape of the TTT plot may now give insight in the underlying lifetime distribution. The following qualitative interpretation of the TTT-plot could be used:

- A plot around the diagonal indicates a constant failure rate, i.e. failure times can be considered exponentially distributed.
- A concave plot (above the diagonal) indicates an increasing failure rate (IFR). A convex plot (under the diagonal) indicates a decreasing failure rate (DFR).

- A plot which fist is convex, and then concave indicates a bathtub like failure rate
- A plot which first is concave, and then convex indicates heterogeneity in the data, see Vatn (1996).

The calculation procedure for obtaining the TTT-plot is shown in Table 23, and the plot is shown in Figure 63.

Table 23 TTT-estimate calculated in EXCEL

i	$T_{(i)}$	$\sum_{j=1}^{i} T_{(j)}$	$\sum_{j=1}^{i} T_{(j)} + (n-i)T_{(i)}$	$\frac{i}{n}$	$\frac{\mathrm{TTT}(T_{(i)})}{\mathrm{TTT}(T_{(n)})}$
1	6.3	6.3	63	0.1	0.06
2	11	17.3	105.3	0.2	0.10
3	21.5	38.8	189.3	0.3	0.18
4	48.4	87.2	377.6	0.4	0.36
5	90.1	177.3	627.8	0.5	0.59
6	120.2	297.5	778.3	0.6	0.73
7	163	460.5	949.5	0.7	0.90
8	182.5	643	1008	0.8	0.95
9	198	841	1039	0.9	0.98
10	219	1060	1060	1	1.00

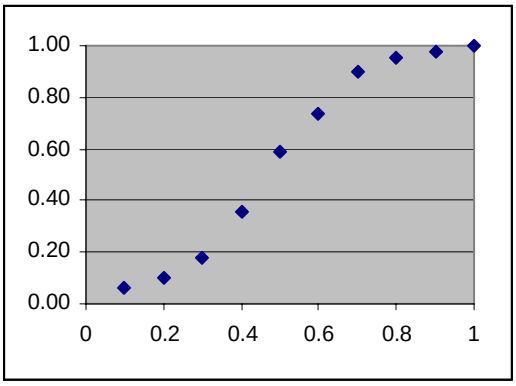


Figure 63 TTT-plot for the example data

14.5.4TTT-plot with TTT-transform as overlay curve

The TTT plot is a non-parametric plot that indicates whether the hazard rate is increasing or not. For parametric distributions it is possible to construct a corresponding parametric curve. This curve is denoted the TTT-transform and is given by:

$$\varphi_F(v) = \frac{1}{MTTF}\int_0^{F^{-1}(v)} (1 - F(u))du \qquad (102)$$

For the Weibull distribution, which is of main interest related to maintenance optimisation, it is shown in e.g. Rausand and Høyland (2003) that the TTT-transform is given by $\varphi_W(v;\alpha)$ = CDFGamma(-ln(1-v),α^{-1},1) where CDFGamma(x, a, b) is the cumulative distribution

function[8] of the gamma distribution with parameters a and b. If we add the TTT-transform to the non-parametric TTT-plot we could vary the parameter α in $\varphi_W(v;\alpha)$ and estimate the ageing parameter α by the value that gives the best fit to the data. Realising that MTTF is estimated by the total service time divided by the number of failures we also easily obtain an estimate for MTTF.

14.5.5More about the ageing parameter

If there is an underlying heterogeneity in the data used for obtaining the ageing parameter α, it could be shown, see e.g. Vatn (1996), that we will underestimate the shape parameter. Thus, if we have estimated the shape parameter by e.g. the TTT plotting technique, and we believe that there is an underlying variation in the data set, we could adjust the estimate for the ageing parameter. Based on the result in Vatn (1996) we could read the adjustment from Figure 64. For example if we have an estimate of the ageing parameter $\alpha = 3$, and assuming medium variation[9], we adjust the estimate to $\alpha = 4.1$.

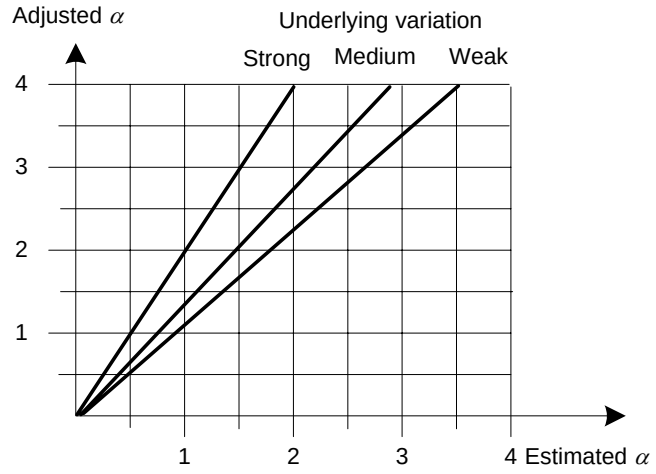


Figure 64 Adjusting the estimate for the shape parameter

Since estimation of the ageing parameter requires a high precision in the collection and analysis of data, we would in some situation use a rather pragmatic approach to reveal the ageing parameter. Based on a systematic qualitative analysis of failure causes and mechanisms, we may use the following "rule of thumb" for assessing the ageing parameter:

$\alpha = 4$: There is a systematic reporting of one and only one particular failure cause or mechanism which is related to ageing, e.g. wear, corrosion, fatigue etc.

$\alpha = 3$: There is a systematic reporting of different failure causes or mechanisms which all are related to ageing, e.g. wear, corrosion, fatigue etc.

$\alpha = 2$: There is a reporting of a mixture of failure causes, some related to ageing, and some not.

$\alpha = 1.5$: Ageing is hardly reported as a failure mechanism.

[8] In MS-EXCEL CDFGamma() is given by the function Gammadist()

[9] The coefficient of variance, CV, is formally used as a measure of variation to produce the results in Figure 64. With "strong" we here mean $CV = 1$, and with "weak" we mean $CV = 0.5$. CV is defined as the mean value divided by the standard deviation.

14.6 COUNTING PROCESS MODELS

14.6.1 Objective

In a counting process model failures are assumed to occur along the time axis, and no assumption is made regarding the status of the unit after the repair is completed. The main objective of the analysis is to reveal any trend in time, and the Nelson Aalen plot is an efficient tool.

14.6.2 Conceptual framework for counting process models

Consider one unit installed at time $t = 0$, observed over a period of time from a to b.

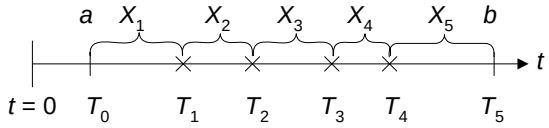


Figure 65 Conceptual model for a counting process

The recorded failure times (global or calendar time) are denoted $T_1, T_2, .., T_n$. By definition $T_0 = a$. The unit is repaired after each failure, but no assumption is made about the quality of the repair. Repair times are considered neglectable. Two extremes are often considered:

- Perfect repair in which case the unit is considered "as good as new" after each repair. In this situation it is reasonable to believe in a Renewal Process (RP), and the theory of life data analysis applies. In Figure 65, the X_i's can be considered as the data set.
- Minimal repair in which case the unit is considered "as bad as old", i.e. the status of the component immediately before the failure occurred. In this situation it is reasonable to believe in a Non-Homogeneous Poison Process (NHPP).

The times between each pair of failures, $X_i = T_i - T_{i-1}$ are denoted the inter-arrival times. If the inter-arrival times tend to become shorter, the system is *deteriorating*. On the other hand, the system is improving if the inter-arrival times tend to become longer (reliability growth).

Note that any trend may be caused to both internal and external circumstances. Typical causes for improving systems are:

- Latent failures are revealed, and fixed
- Improved "organisational environment" due to gained experience of the maintenance and operational personnel
- Improved external environmental conditions
- Failed parts are replaced with new parts with higher reliability

Causes for deteriorating systems are:

- Wear-out mechanisms (of the parts)
- Aggravated external environmental conditions (e.g. more sand in the oil)
- Less resources to maintenance

14.6.3 Nelson Aalen plot

To reveal trends, the Nelson-Aalen plot is constructed. The Nelson-Aalen plot shows the cumulative number of failures on the Y-axis, and the X-axis represents the time. A convex plot indicates a deteriorating system, whereas a concave plot indicates an improving system. The

idea behind the Nelson-Aalen plot is to plot the cumulative number of failures against time. We recall that the ROCOF, $w(t)$, is the failure intensity, and $W(t)$ is the expected cumulative numbers of failures in a time interval:

$$W(t) = \int_0^t w(u)\mathrm{d}u = \mathrm{E}\left[\#\,\text{failures in the interval}\,[0,t)\right] \tag{103}$$

When estimating $W(t)$ we need failure data from one or more processes (systems). Each process (system) is observed in a time interval $(a_i,b_i]$ and t_{ij} denotes failure time j in process i (global or calendar time). The information could be systemised as in Table 24.

Table 24 Example of data for the construction of the Nelson Aalen plot

a_i	b_i	t_{ij}
0	50	7, 20, 35, 44
20	60	26, 33, 41, 48, 57
40	100	50, 60, 69, 83, 88, 92, 99

In order to construct Nelson Aalen plot the following algorithm could be used:

1. Group all the t_{ij}'s in Table 24, sort them, and denote the result t_k, k = 1,2,…..
2. For each k, let O_k denote the number of processes that are under observation just before time t_k
3. Let $\hat{W}_0 = 0$
4. Let $\hat{W}_k = \hat{W}_{k-1} + 1/O_k$, k = 1,2,…
5. Plot $(t_k, \hat{W}_k)$

Note that O_k is the number of processes that are under observation just prior to time t_k, which means that the "jumps" in the estimated cumulative intensity is "adjusted" for the number of processes under observation. The points will follow a straight line if the intensity is constant. If the intensity is increasing,, the t_k's will occur more and more frequent, and the cumulative plot will bend upwards (convex). If the intensity is decreasing the t_k's will occur less and less frequent, and the cumulative plot will bend downwards (concave). Figure 66 shows the Nelson-Aalen plot for the example data in Table 24.

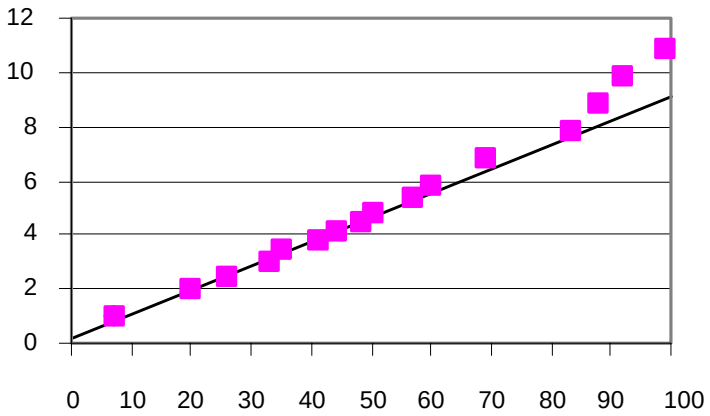


Figure 66 Nelson-Aalen plot for the example data

14.7 Bayesian reliability data analysis

In the previous section we have presented the "frequentiest" or "classical" approach to data analysis. The basic idea up to now has been: The "nature" has provided us with equipment with "true", but unknown reliability parameters. By observing the nature, i.e. counting failure and so on, we try to "reveal" the nature. In the Bayesian framework, however, there exists no "true" reliability parameters. Based on our knowledge, experience and explicit analysis of reliability data, we may state our believes about reliability parameters. We do that in terms of probability statements. These probabilities are, however, not a property of the "nature", but a measure of our knowledge about the system under consideration. We still use the notation of $\boldsymbol{\theta}$ as the (vector) of reliability parameters of interest. The Bayesian approach comprised four basic steps:

1. Specification of a *prior* uncertainty distribution of the reliability parameter, $\pi(\boldsymbol{\theta})$.
2. Structuring reliability data information into a likelihood function, $L(\boldsymbol{\theta};\mathbf{t})$, see Equation (101).
3. Calculation of the *posterior* uncertainty distribution of the reliability parameter vector, $\pi(\boldsymbol{\theta}|\mathbf{t})$.
4. Choosing the Bayes estimate for the reliability parameter, usually the posterior mean.

14.7.1 Specification of prior

The specification of the prior distribution implies that prior to observing the system of interest, we state our believes about the reliability performance of the system. In order to accomplish this we could interview experts, look at data for similar systems and make a statement based on the gathered "information". There exists several formalised procedures for how to perform such "expert judgements" for elicitation of prior distributions, see e.g. Øien and Hokstad (1998). In section 14.4.3 we described a procedure for estimation of a constant failure rate in the "multi sample" situation. The result from such an estimation procedure could also be used to establish an *empirical* prior distribution of the failure rate. The situation is that we have data from k systems that have some similarities with "our" new system. Some of these old systems are "good", and some are "bad". We believe that the reliability performance of the new system is spanned by the reliability performance of these old systems, i.e. the mean θ^* and the variance $\hat{\sigma}^2$ are taken as mean and variance in the prior distribution.

We also need to choose a parametric distribution for the prior. Usually we choose a distribution that is mathematical convenient wrt updating to the posterior distribution. In this presentation we will use the gamma distribution as a prior when we estimate the (constant) failure rate (λ), and the inverted gamma distribution when we estimate the mean time to failure (ξ=MTTF).

Table 25 Prior distributions with characteristics

↓Characteristics Distribution→	Gamma distribution	Inverted gamma distribution
Variable (argument)	λ = failure rate	ξ = MTTF
Probability density function	$\pi(\lambda) \propto \lambda^{\alpha-1}e^{-\beta\lambda}$	$\pi(\xi) \propto (1/\xi)^{\alpha+1}e^{-\beta/\xi}$
Expectation (E)	$E = \alpha/\beta$	$E = \beta/(\alpha\text{-}1)$
Variance (V)	$V = \alpha/\beta^2$	$V = \beta^2(\alpha\text{-}1)^{-2}(\alpha\text{-}2)^{-1}$
Parameter 1	$\alpha = \beta E$	$\alpha = E^2/V + 2$
Parameter 2	$\beta = E/V$	$\beta = E(\alpha\text{-}1)$

For a more comprehensive list of prior distribution candidates, please see e.g. Martz and Waller (1982).

14.7.2The likelihood function

The likelihood function could be seen as a function describing how "likely" the data is wrt the parameter vector. The parameter vector $\boldsymbol{\theta}$ is the argument in the likelihood function, $L(\boldsymbol{\theta};\mathbf{t})$, whereas $\mathbf{t}$ is the observations (data points). When using the likelihood function to update the prior distribution, the data points are seen as fixed numbers. As an example, assume that a failure time, T_1, is exponentially distributed with parameter λ. Given an observation, say t_1, the likelihood function is equal to the probability density function at the value $t = t_1$, but we now treat the parameter λ as the argument, i.e. $L(\lambda,t_1) = f_{T_1}(t_1 \mid \lambda) = \lambda e^{-\lambda t_1}$.

14.7.3Calculating the posterior distribution

The posterior uncertainty distribution of the parameter vector $\boldsymbol{\theta}$, is given by

$$\pi(\boldsymbol{\theta}|\mathbf{t}) \propto L(\boldsymbol{\theta};\mathbf{t}) \times \pi(\boldsymbol{\theta}) \tag{104}$$

Note that $\pi(\boldsymbol{\theta}|\mathbf{t})$ is a probability density function over $\boldsymbol{\theta}$-values, and the proportionality constant should be chosen so that $\pi(\boldsymbol{\theta}|\mathbf{t})$ integrates to one. Usually we do not need to deal with the proportionality constant because $L(\boldsymbol{\theta};\mathbf{t}) \times \pi(\boldsymbol{\theta})$ is recognized as the essential parts of a probability density function.

14.7.4Point estimate and credibility interval for the parameter vector

The posterior uncertainty distribution, $\pi(\boldsymbol{\theta}|\mathbf{t})$ is our believe about the parameter vector $\boldsymbol{\theta}$. In some situations we would like to make a point estimate of the parameter vector $\boldsymbol{\theta}$. Under quadratic loss[10], it could be shown that the Bayes point estimate is given as the posterior mean. We could also state a 100(1-ε)% credibility interval for the parameter vector $\boldsymbol{\theta}$ based on $\pi(\boldsymbol{\theta}|\mathbf{t})$. If $\boldsymbol{\theta}$ is one dimensional, this could easily be accomplished by choosing the interval limits as the $\varepsilon/2$ lower and $\varepsilon/2$ upper percentiles in the posterior distribution.

Example 14.6 – Bayesian failure rate estimate

Assume that we express our prior believe about the failure rate λ of a certain detector type, in terms of the mean value $E = 0.7\times10^{-6}$ (failures per hour), and the standard deviation SD =

[10] Formally we introduce a loss function. This function states that there is a "loss" associated with choosing a too high, or a too low value.

0.3×10^{-6}. Since we are dealing with the failure rate, we choose a gamma distribution ($\pi(\lambda) \propto \lambda^{\alpha-1}e^{-\beta\lambda}$) and from Table 25 we obtain:

$$\beta = E/V = E/SD^2 = (0.7\times10^{-6})/(0.3\times10^{-6})^2 = 7.78\times10^6$$
$$\alpha = \beta E = (7.78\times10^6) \times (0.7\times10^{-6}) = 5.44$$

To establish the likelihood function, we look at the data. In this situation we assume that we have observed identical units for a total time in service, *t,* equal to 525 600 hours (e.g. 60 detector years). In this period we have observed $n = 1$ failure. If we assume exponentially distributed failure times, we know that the number of failures in a period of length *t, N*(*t*), is Poisson distributed with parameter $\lambda\times t$. The probability of observing *n* failures is thus given by:

$$L(\lambda;n,t) = P(N(t) = n) \propto \lambda^n e^{-\lambda\times t} \tag{105}$$

and we have an expression for the likelihood function $L(\lambda;n,t)$.

The posterior distribution is found by multiplying the prior distribution with the likelihood function:

$$\pi(\lambda|n) \propto L(\lambda;n,t) \times \pi(\lambda) \propto \lambda^n e^{-\lambda\times t} \times \lambda^{\alpha-1}e^{-\beta\lambda} \propto \lambda^{(\alpha+n)-1}e^{-(\beta+t)\lambda} \tag{106}$$

and we recognize the posterior distribution as a gamma distribution with new parameters $\alpha' = \alpha + n$, and $\beta' = \beta + t$. The Bayes estimate is given by the mean in this distribution, i.e.

$$\hat{\lambda} = \frac{\alpha+n}{\beta+t} = \frac{5.44+1}{7.78\times10^6+525600} = 0.78\times10^{-6} \tag{107}$$

We note that the maximum likelihood estimator in Equation (97) gives a much higher failure rate estimate (1.9×10^{-6}), but the "weighing procedure" favors the prior distribution in our example. Generally we could interpret α and β here as "number of failures" and "time in service" respectively for the "prior information". □

Exercise 23

Show that if we are working with ξ = MTTF, and assigning an inverted gamma distribution as a prior, then the posterior distribution will also be inverted gamma with parameters $\alpha' = \alpha + n$, and $\beta' = \beta + t$, where *n* is the number of failures observed and *t* is the observation period. □

Table 26 Summary for failure rate and MTTF estimation

↓Characteristics	Failure rate (λ),	MTTF (ξ)
Variable (argument)	λ = failure rate	ξ = MTTF
Prior	$\pi(\lambda)$ ~ Gamma (α,β)	$\pi(\xi)$ ~ Inverted gamma (α,β)
Parameter 1	$\alpha = \beta E$	$\alpha = E^2/V + 2$
Parameter 2	$\beta = E/V$	$\beta = E(\alpha-1)$
Observed # of failures	*n*	*n*
Observation period	*t*	*T*
Posterior	$\pi(\lambda\|n)$ ~ Gamma ($\alpha+n,\beta+t$)	$\pi(\xi\|n)$ ~ Inverted gamma ($\alpha+n,\beta+t$)
Bayes estimate	$\hat{\lambda} = \frac{\alpha+n}{\beta+t}$	$MTTF = \hat{\xi} = \frac{\beta+t}{\alpha+n-1}$

Table 27 Percentage Points of the Chi-square (χ^2) Distribution

$\Pr(Z > z_{\alpha,\nu}) = \alpha$

ν/α	0.995	0.990	0.975	0.950	0.05	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	11.07	12.38	15.09	16.75
6	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00
25	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	124.34	129.56	135.81	140.17

15. FAILURE MODE AND EFFECT ANALYSIS

15.1 Introduction

Failure Mode and Effects Analysis (FMEA) was one of the first systematic techniques for failure analysis. It was developed by reliability engineers in the late 1950's to determine problems that could arise from malfunctions of military systems.

A Failure Mode and Effects Analysis is often the first step in a systems reliability study. It involves reviewing as many components, assemblies and subsystems as possible to identify possible failure modes and the causes and effects of such failures. For each component, the failure modes and their resulting effects on the rest of the system are written onto a specific FMEA form. There are numerous variations of such forms. An example of an FMEA form is shown in Figure 67. The FMEA analysis is an important part of an RCM analysis discussed in Chapter 10. When an FMEA is used as a part of RCM the columns in Figure 67 will be modified to put focus on maintenance issues.

FMECA

System: Performed by:
Subsystem Date:
Function Page

DESCRIPTION OF UNIT			DESCRIPTION OF FAILURE			EFFECT OF FAILURE			FAILURE RATE	CRITICALITY	CORRECTIVE ACTION	REMARKS
IDENTI-FICATION	OPERATIONAL MODE	FUNCTION	FAILURE MODE	FAILURE MECHANISM	HOW TO DETECT	LOCAL	SYSTEM	OPERAT. STATUS				

Figure 67 Example of an FMEA form

A Failure Mode and Effects Analysis is mainly a qualitative analysis, which is usually carried out during the design stage of a system. The purpose is then to identify design areas where improvements are needed to meet the reliability requirements.

The Failure Mode and Effect Analysis can be carried out either by starting at the component level and expanding upwards (the "bottom-up" approach), or from the system level downwards (the "top-down" approach). The component level to which the analysis should be conducted is often a problem to define. It is often necessary to make compromises since the workload could be tremendous even for a system of moderate size. It is, however, a general rule to expand the analysis down to a level at which failure rate estimates are available or can be obtained.

Most Failure Mode and Effects Analyses are carried out according to the "bottom-up" approach. One may, however, for some particular systems save a considerable amount of effort by adopting the "top-down" approach. With this approach, the analysis is carried out in two or more stages. The first stage is an analysis on the functional block diagram level. The possible failure modes and failure effects of each functional block are identified based on knowledge of the block's required function, or from experience on similar equipment. One then proceeds to the next stage, where the components within each functional block are analysed. If a functional block has no failure modes which are critical, then no further analysis of that block needs to be performed. By this screening, it is possible to save time and effort. A weakness of this "top-down" approach lies in the fact that it is not possible to ensure that all failure modes of a functional block have been identified.

An FMEA becomes a Failure Modes, Effects and Criticality Analysis (FMECA) if criticality's or priorities are assigned to the failure mode effects.

The FMEA technique is used as an integral part of an RCM (Reliability Centred Maintenance) analysis. One main idea of RCM is to prevent failures by eliminate or reduce the failure causes. The FMEA analysis should therefore focus on the failure causes and failure mechanisms. When the failure causes and failure mechanisms are identified for each failure mode, it will be possible to suggest time based preventive maintenance actions, or condition monitoring techniques to reduce the resulting failure rate. The proposed maintenance actions are further analysed by means of a so-called RCM logic, and the cost-efficiency are also considered during the RCM analysis.

More detailed information on how to conduct a Failure Mode and Effects Analysis (and an FMECA) may be found in:

- IEC standard 60812 (1985)
- MIL-STD-1629A (1980)
- SAE ARP 926 (1979)

15.2 Structuring

When the FMECA analysis is used as an integral part of an RCM analysis, it is important to clarify the hierarchical structure before one starts filling out the FMECA forms. Since RCM takes a functional approach, the FMECA will also take a "top-down" approach as discussed above. The total analysis will contain three main parts:

- The functional failure analysis
- The completion of the FMECA forms
- The assignment of maintenance tasks

15.3 Elements of functional failure analysis

In principal, we should conduct some formalised functional failure analysis, see Section 10.3. However, due to the huge amount of systems to analyse, the functional analysis is often conducted as a brainstorming process, where the following information is systemised and used as a starting point for the explicit completion of the FMECA forms:

Function name

The function name reflects the functions to be carried out on a relatively high level in the system. In principal we should explicit formulate the function(s) to be carried out. However, often we specify the *equipment class* performing the function. For example "Departure light signal" is specified rather than the more correct formulation "Ensure correct departure light signal".

Description

A description of the function, or equipment class would be appropriate in order to give more information. This could e.g. be list of relevant manufactures, models etc.

Failure modes
For each function, we list relevant failure modes. A failure mode is a description of how the failure manifests seen from the outside. Examples of failure modes for the "Departure light signal" are:

- Wrong signal picture
- Missing signal picture
- Unclear signal picture
- Do not prevent contact hazard in case of earth fault

We observe that the last failure mode in fact is not a failure mode for the "correct" functional description (Ensure correct departure light signal), but is related to another function of the physical "Departure light signal". Thus, if we use a equipment class description rather than an explicit functional statement, the list of failure modes should cover all (implicit) functions of the equipment class.

At the failure mode level, it is also convenient to specify whether the failure mode is evident or hidden, see Figure 68 where we have introduced a "E/H" column.

List of maintenance significant items (MSI)
For each function we also list the relevant items that are required to perform the function. These items will form "rows" in the FMECA forms. Example of maintenance significant items are:

- Signal mast
- Brands
- Background shade
- Earth conductor
- Signal lantern
- Lamp
- Lens
- Transformer

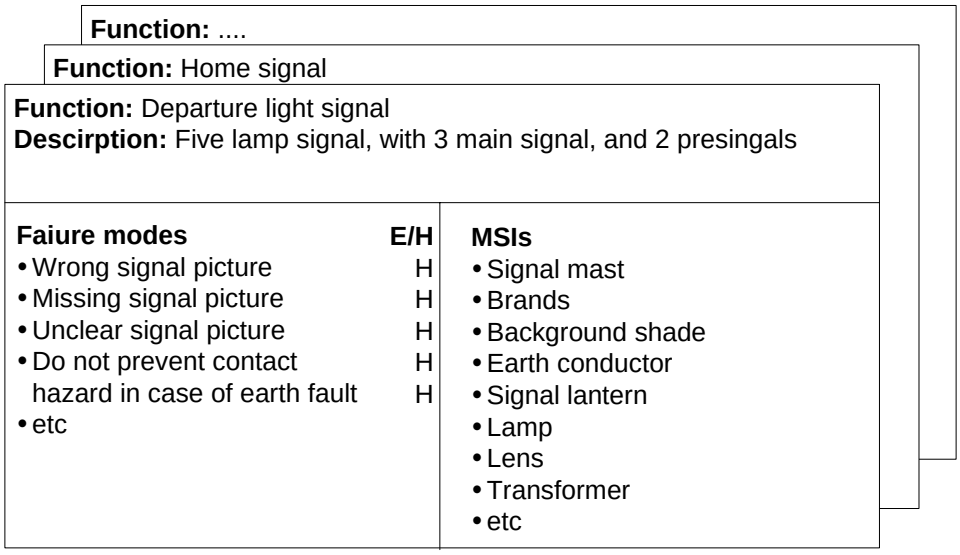


Figure 68 Structure of functional failure analysis

The information entered for the functional analysis could be systematised as indicated in Figure 68.

15.4 Proposed fields for the FMECA forms

In the following a list of fields (columns) for the FMECA forms is proposed. Basically the structure is hierarchical, but the information is presented in a tabular form. The starting point in the FMECA analysis will be the failure modes from the functional failure analysis in section 15.3. Then each maintenance item is analysed with respect to any impact on the various failure modes. In the following we describe the various columns.

Failure mode (equipment class level)
The first column in the FMECA form is the failure mode at the equipment class level identified in the functional failure analysis in section 15.3.

Maintenance significant item (MSI)
The relevant MSI were identified in the functional failure analysis.

MSI function
For each MSI, the functions of the MSI with respect to the current equipment class failure mode are identified.

Failure mode (MSI level)
For the MSI functions we also identify the failure modes at the MSI level. A failure mode is the manner by which a failure is observed, and is defined as non–fulfillment of one of the MSI functions.

Detection method
The detection method column describes how the MSI failure mode could be detected, e.g. by visual inspection, condition monitoring, by the central train control (CTC) system etc.

Hidden or evident
Specify whether the MSI function is hidden or evident.

Demand rate for hidden function, f_D
For MSI functions that are hidden the rate of demand of this function should be specified.

Failure cause
For each failure mode there is one or more failure causes. An failure mode will typically be caused by one or more component failures at a lower level. Note that supporting equipment to the component entered in the FMECA form is for the first time considered at this step. In this context a failure cause may therefore be a failure mode of a supporting equipment. A “no effect” failure of a switch motor may for example be caused by “no electrical current”.

Failure mechanism
For each failure cause, there is one or several failure mechanisms. Examples of failure mechanisms are fatigue, corrosion, and wear. To simplify the analysis, the columns for failure cause and failure mechanism are often grouped into one column.

Mean time between failures
Mean time to failure when no maintenance is performed should be specified. I.e. what would we anticipate the MTTF will be if no preventive maintenance is carried out. The MTTF is specified for one component if it is a “point” object, and for a standardised distance if it is a “line” object such as rails, sleepers etc.

Local effect of failure
The local effect of a failure mode could be effects on other MSIs, or the failure mode on the equipment class level, optionally with a broader description.

Global effect of failure
The global effect of the failure mode usually relates to TOP-level functions, and especially to safety and punctuality issues.

TOP-event safety
The TOP-event in this context is the accidental event that might be the result of the failure mode. Within railway application it is common to define the following seven TOP events:

- Derailment
- Collision train-train
- Collision train-object
- Fire
- Persons injured or killed in or at the track
- Persons injured or killed at level crossings
- Passengers injured or killed at platforms

Barrier against TOP-event safety
This field is used to list barriers that are designed to prevent a failure mode from resulting in the safety TOP-event. For example brands on the signalling pole would help the locomotive driver to recognize the signal in case of a dark lamp.

$P_{TE\text{-}S}$
This field is used to assess the probability that the other barriers against the TOP-event all fails. $P_{TE\text{-}S}$ should count for all the barriers listed under "Barrier against TOP-event safety".

TOP-event puncutality
The following TOP events for punctuality is currently proposed:

- Full stop (Infrastructure)
- Slow speed (Infrastructure)
- Manual train operation – line block (Infrastructure)
- Manual train operation – station (Infrastructure)
- Full stop – First line maintenance (Rolling stock)
- Full stop – Depot maintenance (Rolling stock)
- ATP failure–80 km/h (Rolling stock)
- Slow speed –40 km/h (Rolling stock)

The relation between the TOP-event for punctuality and "Passenger delay minutes" is generally very complex, and a mathematical model is not supported here.

Barrier against TOP-event punctuality
This field is used to list barriers that are designed to prevent a failure mode from resulting in the punctuality TOP-event. Since the fail safe principle is fundamental in railway operation, there are usually no barriers against the punctuality TOP-event when a component fails. Examples of barriers could be 2oo3 voting on some critical components within the system.

$P_{TE\text{-}P}$
This field is used to assess the probability that the other barriers against the punctuality TOP-event all fails. $P_{TE\text{-}P}$ should count for all the barriers listed under "Barrier against TOP-event punctuality". Due to the fail safe principle, $P_{TE\text{-}P}$ will often be equal to one.

Other consequences
Other consequences could also be listed. Some of these are non-quantitative like noise effects, passenger comfort, and aesthetics. Material damages to rolling stock, or components in the infrastructure could also be listed. Material damages could be categorized in terms of monetary values, but this is not pursued here.

Exposure measure
In order to capture the significance of the actual failure mode one have to consider the number of components, or the length of a line object. It would often be convenient to consider a "standardised" track section of e.g. 500 km. For point object, we then list the "average" number of MSIs on such a track, and for line objects we simply give the length, e.g. 500 km.

Mean Down Time (MDT)
The mean down time is the time from a failure occurs until the failure has been fixed and any traffic restrictions has been removed.

Safety criticality
The safety criticality is a measure comprising the following fields:

- MTTF
- $P_{TE\text{-}S}$
- TOP-event safety
- Exposure measure (*EM*)

Formally, we could let the criticality measure reflect the PLL contribution of the actual failure mode if no preventive maintenance is carried out:

$$PLL = MTTF^{-1} \times EM \times P_{TE\text{-}S} \times \sum_{j=1:6}(PC_j \times PLL_j) \quad (108)$$

Where PC_j is the probability that the safety TOP event results in consequence class *Cj* and PLL_j is the PLL contribution of consequence class *Cj* (see Table 5 and Table 6 in section 11.1 page 99. A standardization of PLL classes could be defined. Typically we use the following classes:

- Red: PLL contribution is unacceptable. Preventive maintenance actions are required
- Yellow: PLL contribution is acceptable. Preventive maintenance action should be considered only if it obviously will be cost efficient.
- Green: PLL contribution is low. Preventive maintenance action should be considered only if it obviously will be cost efficient. Normally we will accept a corrective maintenance activity if the failure mode occurs.
- White: PLL contribution is neglect able. Not necessary to consider any preventive maintenance action.

Punctuality criticality
The punctuality criticality is a measure comprising the following fields:

- MTTF
- $P_{TE\text{-}P}$
- TOP-event punctuality
- Exposure measure (*EM*)
- MDT

Since there is a complex relation between delay time minutes and the above parameters it is difficult to establish a good criticality measure for punctuality. A very simple measure for the delay time minutes (DTM) is:

$$DTM = \text{MTTF}^{-1} \times EM \times P_{TE\text{-}P} \times \text{MDT} \times CF \quad (109)$$

Where *CF* is a correction factor. The correction factor should account for the number of trains that would be affected by a failure, and the severity of the TOP-event. For example a full stop is more critical than speed reduction.

A standardization of DTM classes could be defined. Typically we use the following classes:

- Red: DTM contribution is unacceptable. Preventive maintenance actions are required
- Yellow: DTM contribution is acceptable. Preventive maintenance action should be considered only if it obviously will be cost efficient.
- Green: DTM contribution is low. Preventive maintenance action should be considered only if it obviously will be cost efficient. Normally we will accept a corrective maintenance activity if the failure mode occurs.
- White: DTM contribution is neglect able. Not necessary to consider any preventive maintenance action.

15.5 The assignment of maintenance tasks

If a failure mode is considered significant with respect to safety or punctuality (or other dimensions) a preventive maintenance should be assigned. In order do such an assignment, further information has to be specified. This could be done as part of the FMECA form discussed in Section 15.4, or we could establish a separate form. In the following we assume that a special form is used for this part of the analysis. The following fields are recommended:

FMECA ID
A link to the FMECA form should be made for each maintenance task. In principle there could be more than one maintenance task for each MSI failure mode, hence there is a one to many relationship. Note also that one maintenance task could affect several failure modes, hence there is in principle a many to many relationship between the list of MSI failure modes and the maintenance task.

Failure propagation
For each failure cause the failure propagation should be described in terms of categories 1-4 of Figure 20 to Figure 23 in Chapter 6.

Length of PF-interval
The expected value and the standard deviation of the PF interval should be entered when relevant, see Section 7.3.2.

Ageing parameter
For non-observable failure progression ageing effects should be described. Relevant categories are *strong*, *moderate* or *low* ageing effects. As an alternative to specifying the ageing parameter, the safe time to failure (STTF) could be entered, see Section 7.3.2.

Maintenance task
The maintenance task is determined by the RCM logic discussed in Section 10.8 page 93.

Preliminary maintenance interval

A formalised approach is required to optimise the maintenance interval. However, at this stage of the analysis it would be appropriate to specify a preliminary estimate.

16. Hazard and Operability (HAZOP) study

16.1 Introduction

A Hazard and Operability (HAZOP) study is a structured and systematic examination of a planned or existing process or operation in order to identify and evaluate problems that may represent risks to personnel or equipment, or prevent efficient operation.

The HAZOP technique was initially developed to analyse chemical process systems, but has later been extended to other types of systems and also to complex operations and to software systems. With respect to maintenance, the HAZOP method could be applied with the following objective:

- Analysis of the technical system in order to find weak points where a maintenance task could reduce the probability of failure, and/or the consequence of a failure
- Analysis of the maintenance action (procedure HAZOP) where the objective is to identify critical tasks when executing the maintenance.

A HAZOP is a qualitative technique based on guide-words and is carried out by a multi-disciplinary team (HAZOP team) during a set of meetings.

The HAZOP study should preferably be carried out as early in the design phase as possible - to have influence on the design. On the other hand; to carry out a HAZOP we need a rather complete design. As a compromise, the HAZOP is usually carried out as a final check when the detailed design has been completed.

A HAZOP study may also be conducted on an existing facility to identify modifications that should be implemented to reduce risk and operability problems.

16.2 Types of HAZOP

There exist several types of HAZOP, and we often differentiate between the following types:

- Process HAZOP
  - The HAZOP technique was originally developed to assess plants and process systems
- Human HAZOP
  - A “family” of specialized HAZOPs. More focused on human errors than technical failures
- Procedure HAZOP
  - Review of procedures or operational sequences, sometimes denoted SAFOP - SAFe Operation Study
- Software HAZOP
  - Identification of possible errors in the development of software

16.3 The HAZOP procedure

As a basis for the HAZOP study the following information should be available:

- Process flow diagrams
- Piping and instrumentation diagrams (P&IDs)

- Cause and effect (C&E) diagrams
- Layout diagrams
- Material safety data sheets
- Provisional operating instructions
- Heat and material balances
- Equipment data sheets
- Start-up and emergency shut-down procedures

The following steps are often used in a HAZOP procedure

1. Divide the system into sections (i.e., reactor, storage)
2. Choose a study node
3. Describe the design intent
4. Select a process parameter
5. Apply a guide-word
6. Determine cause(s)
7. Evaluate consequences/problems
8. Recommend action: What? When? Who?
9. Record information
10. Repeat procedure (from step 2)

In the following some of the steps are briefly discussed.

A *study node* could be a line, a vessel, a pump, or an operating instruction. When studying a node it might be necessary to consider different operational modes, e.g.

- Normal operation
- Reduced throughput operation
- Routine start-up
- Routine shutdown
- Emergency shutdown
- Commissioning
- Special operating modes

The *design intent* is a description of how the process is expected to behave at the node; this is qualitatively described as an activity (e.g., feed, reaction, sedimentation) and/or quantitatively in the process parameters, like temperature, flow rate, pressure, composition, etc.

A process parameter is a parameter describing the process or the activity being analyzed. Examples of process parameters are shown in Figure 69:

Flow	Composition	pH
Pressure	Addition	Sequence
Temperature	Separation	Signal
Mixing	Time	Start/stop
Stirring	Phase	Operate
Transfer	Speed	Maintain
Level	Particle size	Services
Viscosity	Measure	Communication
Reaction	Control	

Figure 69 HAZOP process parameters

A *guide word* short word to create the imagination of a deviation of the design/process intent. The most commonly used set of guide-words is: no, more, less, as well as, part of, other than, and reverse. In addition, guide-words like too early, too late, instead of, are used; the latter mainly for batch-like processes. The guide-words are applied, in turn, to all the parameters, in order to identify unexpected and yet credible deviations from the design/process intent. HAZOP guide-words are listed in Table 28:

Table 28 HAZOP guide-words

Guide word	Meaning	Example
No (not, none)	None of the design intent is achieved	No flow when production is expected
More (more of, higher)	Quantitative increase in a parameter	Higher temperature than designed
Less (lessof, lower)	Quantitative decrease in a parameter	Lower pressure than normal
As well as (more than)	An additional activity occurs	Other valves closed at the same time
Part of	Only some of the design intention is achieved	Only part of the system is shut down
Reverse	Logical opposite of the design intention	Back-flow when the system shuts down
Other than (other)	Complete substitution – another activity takes place	Liquids in the gas piping
Early / late	The timing is different from the intention	The valve is opened to late
Before / after	The step (or part of it) is effected out of sequence	The work starts before the high voltage is disconnected
Faster / slower	The step is done/not done with the right timing	Oil is removed faster than the sink can swallow
Where else	Applicable for flows, transfer, sources and destinations	The fluid is emptied in the wrong bottle

A guide-word applied to a process parameter gives a deviation

Examples:

- No & Flow
- No flow $\Rightarrow$ dehydration

- More & Flow
  - More flow ⇒ flash flow
- More & Pressure
  - More pressure ⇒ overpressure

A simple example HAZOP worksheet is shown in Table 29:

Table 29 Example of HAZOP worksheet for the process parameter flow

GW	Deviation	Consequences	Causes	Recommend action
No	No flow	Too much ammonia in the reactor. Discharge to working area	1. Valve A fails in closed position 2. Phosphoric acid depot is empty 3. Pipe blockage, or pipe fractured	Automatic closure of valve B when no flow from phosphoric depot
Less	Les flow	Too much ammonia in the reactor. Discharge to working area. Investigate the situation!	1. Valve A partly closed 2. Pipe partly blocked, or fractured	Automatic closure of valve B when flow is missing or is reduced from phosphoric depot Set-point determined by toxicity and flow limitations
More	More flow	Too much phosphoric acid. No danger in working area		

A more comprehensive HAZOP worksheet is shown in Figure 70:

GW	Dev.	Causes	Consequence	Safeguards	S	L	R	Recs	Remarks	Comments

Figure 70 HAZOP worksheet (Nolan 1994)

The columns are as follows:

GW (Guidewords)

Simple word or phrase used to generate deviations (or hazards) associated with a process equipment or process section. Examples: pressure, flow, temperature etc.

Dev. (Deviation)

Deviation from the design or operation intention associated with the guideword (too high, too low, more, less, reverse, etc)

Causes

Reason for hazard or deviation to occur (failures, wrong operation, etc)

Consequence

The effect of a deviation or hazard associated with the causes.

Note that no credit should be given for any safeguard at this stage. For example; Even though a high level alarm would activate a downstream equipment shutdown, the *consequence* of possible liquid carry over and damage to downstream equipment should still be described. The high level alarm should be described as a *safeguard*.

Safeguards

Measures *present in design* to be taken to prevent or mitigate the risk of an accident (operator surveillance, instrumentation, ESD, blow down, etc).

Note that there are some requirements to what can be assigned as safeguards, and one key word is “independence”. If the cause of a hazard is within a control loop, a safeguard should be independent of that control loop, meaning that alarms from the control system should not be assigned as safeguard.

S (Severity of consequences, taken into account present safeguards)

The magnitude of physical or intangible loss consequences.

L (Likelihood or Probability):

The measure of the expected frequency of an event’s occurrence.

R (Ranking or Resulting Risk)

The qualitative estimation of risk from severity and likelihood. The aim is to provide a prioritizing of risk based on its magnitude.

Recs (Recommendations)

Additional activities identified which may reduce the risk by either reducing the severity or likelihood.

Remarks

Other information related to the review (project decisions, related data, pending studies etc).

Comments

Supplemental *technical* information about the equipment or process section discussed.

Note that the HAZOP study could be quite time consuming since each guide word should be applied to all process parameters, and this should be repeated for all of the study nodes.

17. FAULT TREE ANALYSIS

17.1 Introduction

A fault tree is a logic diagram that displays the relationships between a potential critical event (accident) in a system and the reasons for this event. The reasons may be environmental conditions, human errors, normal events (events which are expected to occur during the life span of the system) and specific component failures. A properly constructed fault tree provides a good illustration of the various combinations of failures and other events which can lead to a specified critical event. The fault tree is easy to explain to engineers without prior experience of fault tree analysis.

An advantage with a fault tree analysis is that the analyst is forced to understand the failure possibilities of the system, to a detailed level. A lot of system weaknesses may thus be revealed and corrected during the fault tree construction.

A fault tree is a *static* picture of the combinations of failures and events which can cause the TOP event to occur. Fault tree analysis is thus not a suitable technique for analysing dynamic systems, like switching systems, phased mission systems and systems subject to complex maintenance strategies.

A fault tree analysis may be qualitative, quantitative or both, depending on the objectives of the analysis. Possible results from the analysis may e.g. be:

- A listing of the possible combinations of environmental factors, human errors, normal events and component failures that can result in a critical event in the system.
- The probability that the critical event will occur during a specified time interval.

The analysis of a system by the fault tree technique is normally carried out in five steps:

1. Definition of the problem and the boundary conditions.
2. Construction of the fault tree.
3. Identification of minimal cut and/or path sets.
4. Qualitative analysis of the fault tree.
5. Quantitative analysis of the fault tree.

In the following we will present the basic elements of standard fault tree analysis. Then we will conclude this chapter by presenting a numerical example illustrating how the technique could be utilised in relation to maintenance optimisation.

17.2 Fault tree construction

17.2.1 Fault tree diagram, symbols and logic

A fault tree is a logic diagram that displays the connections between a potential system failure (TOP event) and the reasons for this event. The reasons (Basic events) may be environmental conditions, human errors, normal events and component failures. The graphical symbols used to illustrate these connections are called “logic gates”. The output from a logic gate is determined by the input events.

The graphical layout of the fault tree symbols are dependent on what standard we choose to follow. The table below shows the most commonly used fault tree symbols together with a brief description of their interpretation.

17.2.2Definition of the Problem and the Boundary Conditions

This activity consists of:

- Definition of the critical event (the accident) to be analysed.
- Definition of the boundary conditions for the analysis.

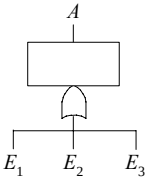
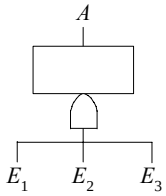
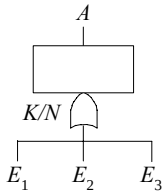
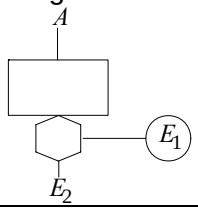
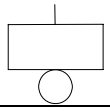
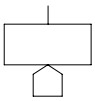
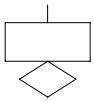
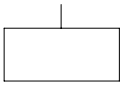
The critical event (accident) to be analysed is normally called the TOP event. It is very important that the TOP event is given a clear and unambiguous definition. If not, the analysis will often be of limited value. As an example, the event description "Fire in the plant" is far too general and vague. The description of the TOP event should always answer the questions: **What, where** and **when**?

> **What:** Describes what type of critical event (accident) is occurring, e.g. collision between two trains.
>
> **Where:** Describes where the critical event occurs, e.g. on a single track section.
>
> **When:** Describes when the critical event occurs, e.g. during normal operation.

A more precise TOP event description is thus: "Collision between two trains on a single track section during normal operation".

- To get a consistent analysis, it is important that the *boundary conditions* for the analysis are carefully defined. By boundary conditions we mean: **The physical boundaries of the system**. What parts of the system are to be included in the analysis, and what parts are not?
- **The initial conditions**. What is the operational state of the system when the TOP event is occurring? Is the system running on full/reduced capacity? Which valves are open/closed, which pumps are functioning etc.?
- **Boundary conditions with respect to external stresses**. What type of external stresses should be included in the analysis? By external stresses we here mean stresses from war, sabotage, earthquake, lightning etc.
- **The level of resolution**. How far down in detail should we go to identify potential reasons for a failed state? Should we as an example be satisfied when we have identified the reason to be a "valve failure", or should we break it further down to failures in the valve housing, valve stem, actuator etc.? When determining the required level of resolution, we should remember that the detail in the fault tree should be comparable to the detail of the information available

Table 30 Fault tree symbols.

	SYMBOL	DESCRIPTION
LOGIC GATES	"OR" gate 	The OR-gate indicates that the output event *A* occurs if any of the input events E_i occurs.
	"AND" gate 	The AND-gate indicates that the output event *A* occurs only when all the input events E_i occurs simultaneously.
	"KooN" gate 	The KooN-gate indicates that the output event *A* occurs if K or more of the input events E_i occurs.
	"Inhibit" gate 	The INHIBIT gate indicates that the output event *A* occurs if both the conditional event E_1 and the input event E_2 occoccur.
INPUT EVENTS	"BASIC" event 	The Basic event represents a basic equipment fault or failure that requires no further development into more basic faults or failures.
	"HOUSE" event 	The House event represents a condition or an event which is TRUE (ON) or FALSE (OFF) (not true).
	"UNDEVELOPED" event 	The Undeveloped event represents a fault event that is not examined further because information is unavailable or because its consequence is insignificant.
DESCRIPTION OF STATE	"COMMENT rectangle 	The Comment rectangle is for supplementary information.
TRANSFER SYMBOLS	"TRANSFER" down  "TRANSFER" up 	The Transfer **down** symbol indicates that the fault tree is developed further at the occurrence of the corresponding Transfer **up** symbol.

17.2.3 Construction of the Fault Tree

The fault tree construction always starts with the TOP event. We must thereafter carefully try to identify all fault events which are the immediate, necessary and sufficient causes that result in the TOP event. These causes are connected to the TOP event via a logic gate. It is important that the first level of causes under the TOP event is developed in a structured way. This first level is often referred to as the TOP structure of the fault tree. The TOP structure causes are often taken to be failures in the prime modules of the system, or in the prime functions of the system. We then proceed, level by level, until all fault events have been developed to the required level of resolution. The analysis is in other words deductive and is carried out by repeated asking "What are the reasons for...?"

Rules for fault tree construction:

- **Description of the fault events**. Each of the Basic events must be carefully described (what, where, when) in a "rectangle".
- **Evaluation of the fault events.** Component failures may be divided in three groups: primary failures, secondary failures and command faults.
  - A primary failure is a failure caused by natural ageing of the component. The primary failure occurs under conditions within the design envelope of the component. A repair action is necessary to return the component to a functioning state.
  - A secondary failure is a failure caused by excessive stresses outside the design envelope of the component. A repair action is necessary to return the component to a functioning state.
  - A command fault is a failure caused by an improper control signal or noise. A repair action is usually not required to return the component to a functioning state. Command faults are often referred to as transient failures.
  - The "normal" Basic events in a fault tree are primary failures identifying the equipment which is responsible for the failure. Secondary failures and command faults are intermediate events which require a further investigation to identify the prime reasons.

  When evaluating a fault event, we ask the question "can this fault be a primary failure?". If the answer is "yes", we classify the fault event as a "normal" Basic event. If the answer is "no", we classify the fault event as either an intermediate event which has to be further developed, or as a "secondary" Basic event. The "secondary" Basic event is often called an "Undeveloped" event and represents a fault event that is not examined further because information is unavailable or because its consequence is insignificant.
- **The gates shall be completed.** All inputs to a specific gate should be completely defined and described before proceeding to the next gate. The fault tree should be completed in levels, and each level should be completed before beginning the next level.

17.3 Identification of Minimal Cut- and Path Sets

A fault tree provides valuable information about possible combinations of fault events which can result in a critical failure (TOP event) of the system. Such a combination of fault events is called a cut set.

*A **cut set** in a fault tree is a set of Basic events whose (simultaneous) occurrence ensures that the TOP event occurs. A cut set is said to be **minimal** if the set cannot be reduced without loosing its status as a cut set.*

*A **path set** in a fault tree is a set of Basic events whose non-occurrence (simultaneously) ensures that the TOP event does not occur. A path set is said to be **minimal** if the set cannot be reduced without loosing its status as a path set.*

For small and simple fault trees, it is feasible to identify the minimal cut- and path sets by inspection without any formal procedure/algorithm. For large or complex fault trees we need an efficient algorithm. The MOCUS algorithm (*Method for obtaining cut sets*) is described in standard FTA textbooks, and an efficient improvement of the algorithm is described by Vatn (1993). We could choose to work with either the minimal cut sets or the minimal path sets. In the following we present an approach based on minimal cut sets.

Exercise 24
Consider the hydro power system in Figure 71.

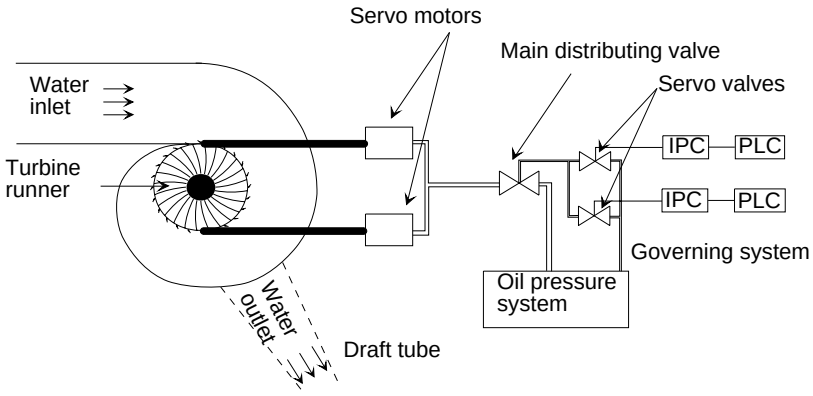


Figure 71 Hydro power turbine with governing system

In order to control the frequency of the turbine runner (TR) both servo motors (SM) have to function. The main distributing valve (MDV) is controlled by two servo valves (SV). Each servo valve is a gain controlled by a programmable logical controller (PLC) via an input card (IPC). It is sufficient that one servo valve with IPC and PLC is functioning in order to have the main distributing valve to operate. The oil pressure system (OPS) comprises both an oil tank, and an oil pump.

a. Define the TOP event by asking the three questions What, Where and When.
b. Establish the fault tree for this system.
c. Find the minimal cut sets by direct inspection of the fault tree (you might alternatively download CARA FaultTree http://www.sydvest.com/Cara-demo/Demo.ASP

□

17.4 Qualitative Evaluation of the Fault Tree

A qualitative evaluation of the fault tree may be carried out on the basis of the minimal cut sets. The importance of a cut set depends obviously on the number of Basic events in the cut set. The number of different Basic events in a minimal cut set is called the *order* of the cut set.

A cut set of order one is usually more critical than a cut set of order two, or higher. When we have a cut set with only one Basic event, the TOP event will occur as soon as this Basic event occurs. When a cut set has two Basic events, both of these have to occur at the same time to cause the TOP event to occur.

Another important factor is the type of Basic events in a minimal cut set. We may rank the criticality of the various cut sets according to the following ranking of the Basic events:

1. Human error
2. Failure of active equipment
3. Failure of passive equipment

The ranking is based on the assumption that human errors occur more frequently than active equipment failures, and that active equipment is more failure-prone than passive equipment (an active or running pump is for example more exposed to failures than a passive standby pump).

17.5 Quantitative Analysis of the Fault Tree

17.5.1 Important system reliability measures

When reliability data for each of the basic events is available, it is possible to carry out a quantitative evaluation of the fault tree. Different system reliability measures may be of interest:

- $Q_0(t)$ - The probability that the TOP event occurs at time *t.*
- $R_0(t)$ - The probability that the TOP event does not occur in [0,*t*).
- MTTF_0 - Mean time to first system failure.
- F_0 - TOP event frequency.

17.5.2 $Q_0(t)$ - The probability that the TOP event occurs at time *t*

$Q_0(t)$ is the probability that the TOP event is occurring at time *t*. If the state of each component[1)] in the fault tree is known at time *t*, then the state of the TOP event can also be determined regardless of what has happened up to time *t*. Hence $Q_0(t)$ is uniquely determined by the component unavailabilities, i.e. the $q_i(t)$'s.

If all components have failure data of the category[1)] *on demand probability*, the $q_i(t)$'s are constant with respect to the time, hence $Q_0(t)$ is also time invariant. If at least one component in each minimal cut set has data of the category *repairable unit* or *non-repairable unit*, the corresponding $q_i(t)$'s will increase from $q_i(0) = 0$ to some asymptotic value $q_i(\infty) \leq 1$ implying $Q_0(t)$ to increase from $Q_0(0) = 0$ to $Q_0(\infty) \leq 1$.

It makes no sense to obtain values for $Q_0(t)$ when components with failure data of category *frequency* is used. Components with failure data of category *frequency* are assumed to function at time *t* with probability one (*duration* of occurrence equals zero). Thus minimal cut sets with such components are also assumed to function at time *t* with probability one.

---

11) We will use the term *component* instead of *input event* because it is natural to think about the occurrence of an input event as a component failure. In other situations, e.g. when the input event represent a *human error*, this is not natural.

12) The failure data categories are defined in Section 17.6.

17.5.3$R_0(t)$ - The probability that the TOP event does not occur in [0,t)

$R_0(t)$ is the probability that the TOP event has *not* occurred in the time period from 0 to t, i.e. the probability that the system has survived up to time t.

In opposition to $Q_0(t)$, $R_0(t)$ *does* depend on what has happened up to time t, and not only the situation at time t. We will illustrate this by considering a system with two components A and B in parallel. This corresponds to two components connected with an AND-gate. The TOP event is occurring if both A and B are occurring at time t, hence

$$Q_0(t) = q_A(t) \times q_B(t) \tag{110}$$

To determine whether the TOP event does occur one or several times up to time t, it is not sufficient to know that both components have failed one or several times up to time t. This because the TOP event will *not* occur if one of the component is functioning while the other is repaired.

As a special case, when *all* components have failure data of category *non-repairable unit*, we have

$$R_0(t) = 1 - Q_0(t) \tag{111}$$

Generally Monte Carlo techniques or use of numerical integration is requited to calculate $R_0(t)$.

17.5.4$MTTF_0$ - Mean time to first system failure

$MTTF_0$ is the mean time to the *first* failure of the TOP event. The $MTTF_0$ is always greater or equal to the mean time *between* failures, MTBF. This is because all components are assumed to function at time t, but this assumption can not be made when the system has been restored after a system failure. Generally Monte Carlo techniques or use of numerical integration is requited to calculate $MTTF_0$.

17.5.5F_0 – *Frequency of TOP event*

The frequency of the TOP event is the expected number of occurrences of the TOP event in a period of time, for example:

$$F_0 = 2 \text{ occurrences per year} \tag{112}$$

Note that the number of occurrences of the TOP event, say X, in a given period of time, is a *random number*. We may be interested in obtaining the *distribution* of X as well as the *expected value* of X, $E(X)$. In this presentation we always interpret F_0 as the expected number of occurrence of the TOP event during a time period.

A common situation when the frequency of the TOP event applies, is when one and only one component in each minimal cut set has failure data of category *frequency*. As an example, consider a system with two components A and B in parallel. Component A has data of failure category *frequency*, say f_A, and component B has failure data of category *on demand probability*, say q_B. We then have:

$$F_0 = f_A \cdot q_B \tag{113}$$

This will be a typical situation when A is an undesired event and B is a barrier.

17.5.6 Notations for describing reliability measures

We will end this chapter by giving an overview of the notation used when describing reliability measures. The overview is given in Table 31.

Table 31 Summary of FTA notation

Notation	Description
$Q_0(t)$	P(the TOP event occurs at time t).
$\overset{\vee}{Q}_j(t)$	P(cut set j occurs at time t)
$R_0(t)$	P(the TOP event does not occur in [0,t)).
MTTF	Mean time to first system failure
F_0	Frequency of the TOP event
$q_i(t)$	P(i'th component is not functioning at time t)
λ_i	Failure rate, i'th component, i.e. expected number of failures of i'th component per hours
f_i	Frequency of i'th input event i.e. expected number of occurrences of i'th input event per hours
MDT_i	Mean down time, MDT, for i'th component (in hours)
τ	Length of test interval for components periodically tested (in hours)
$I^B(i\|t)$	Birnbaum's Measure of Reliability Importance for component i
$I^{VF}(i\|t)$	Vesely-Fussell's Measure of Reliability Importance for component I
$I^{IP}(i\|t)$	Improvement Potential Reliability Measure for component I
$I^{CR}(i\|t)$	Criticality Importance Reliability Measure for component I
$I^O(i)$	Order of smallest cut set for component I
$B_\phi(i)$	Birnbaum's Measure of Structural Importance for component I
$I^{CI}(j)$	Cut set importance of cut set j

17.6 Input Data to the Fault Tree

17.6.1 Category of failure data for input events

The crucial factors in the quantitative evaluation of the fault tree are the reliability data for the input events. Table 32 lists five different categories of failure data for input events that often are relevant:

Table 32 Category of failure data for Input events

Category of failure data	Reliability Parameters
Frequency	f = Frequency [1)]
On demand probability	q = Probability
Test interval	τ =Test interval [2)], MDT = Mean Down Time [2)] and λ = Failure rate[3)]
Repairable unit	MDT = Mean Down Time [2)] and λ = Failure rate[3)]
Non repairable unit	λ = Failure rate[3)]

1) Expected number of occurrences per time unit, e.g. hours.
2) To be specified according to the chosen time unit.
3) Expected number of failures per time unit.

17.6.2Frequency

This category is used to describe events occurring now and then, but with no duration. Thus the *probability* that the event is occurring at time t, $q_i(t) = 0$.

Note! If there is a *duration* of the event, the event should be described as a *repairable unit*, where the failure rate equals the frequency of the event, and the mean down time equals the duration.

17.6.3On demand probability

This category is usually used to describe components which is *not* activated during normal operation. The component is demanded only now and then. The reliability data represents the probability that the component is not able to perform its function upon request. In safety systems, the *operator* is often modelled by an *on demand probability*, for example: *Operator fails to activate manual shut-down system*.

17.6.4Test interval

This category is used to describe components which are tested periodically with test interval τ. A failure may occur anywhere in the test interval. The failure will, however, not be detected until the test is carried out or the component is needed. This is a typical situation for many types of detectors, process sensors and safety valves. The probability $q_i(t)$ is in this situation often referred to as the mean fractional dead-time, MFDT. The reliability parameters entered are the failure rate λ, the test interval τ (in hours) and the mean down time MDT. The probability of failure on demand (PFD) may be approximated by the formula:

$$q_i \approx \frac{\lambda\tau}{2} + \lambda \mathrm{MDT} \tag{114}$$

Note that this formula only is valid if we have *independent* testing of each component. If components are tested *simultaneously*, or if we have *staggered* testing, this formula will not be correct.

17.6.5Repairable unit

The component is repaired when a failure occurs. If the failure rate is denoted λ and the mean time to repair (MTTR) is denoted τ, and $q_i(t)$ may be calculated by the formula:

$$q_i(t) = \frac{\lambda \times \mathrm{MDT}}{1 + \lambda \times \mathrm{MDT}} \left(1 - e^{-\frac{(1+\lambda \times \mathrm{MDT})t}{\mathrm{MDT}}} \right) \tag{115}$$

By letting t tend to infinity, we obtain the well-known approximation:

$$q_i(t) = \frac{\lambda \cdot \mathrm{MDT}}{1 + \lambda \cdot \mathrm{MDT}} = \frac{\mathrm{MDT}}{\mathrm{MDT} + \mathrm{MTTF}} \tag{116}$$

where

$$MTTF = \frac{1}{\lambda} \tag{117}$$

17.6.6Non repairable unit

The component is not repaired when a failure occurs. If the failure rate of the component is denoted by λ, then:

$$q_i(t) = 1 - e^{-\lambda t} \tag{118}$$

17.7 TOP Event Calculations

We will now describe simple approximation formulas for the following system measures:

- Q_0 - The probability that the TOP event occurs.
- F_0 - TOP event frequency.
- $MTTF_0$ - Mean time to first system failure.

Note that we in the following drop the time index *t* in order to keep the presentation as simple as possible. The starting point for the calculations will be the minimal cut sets, and reliability figures for each basic event. Here it is sufficient to consider the probability (*q*) of a basic event occurrence and the frequency (*f*) of basic event occurrence.

For a thorough presentation of formulas and calculation methods we refer to standard text books in reliability theory, e.g. Rausand and Høyland (2004).

17.7.1Q_0 – The TOP event probability

The TOP event probability Q_0 depends on the structure of the fault tree (minimal cut sets) and the probabilities that the various basic events occurs. In order to calculate Q_0 we use an approximation formula. The idea is to sum the contribution from each cut set. Let the minimal cut sets be denoted $K_1,K_2,...,K_k$, and assume that the basic events are independent. Then the probability that minimal cut set K_j occurs is given by the product of the basic event occurrence probabilities:

$$\breve{Q}_j = \prod_{i \in K_j} q_i \tag{119}$$

Summing the contributions gives:

$$Q_0 \approx \breve{Q}_1 + \breve{Q}_2 + ... + \breve{Q}_k = \sum_{j=1}^{k} \breve{Q}_j \tag{120}$$

Generally some minimal cut sets will have common basic events, and hence the expression above is an approximation. The approximation is, however, an upper limit, and the approximation is good when the q_i's are close to 0 (<0.01), Figure 72 shows a graphical illustration of the procedure. In this example we assume that we have the following minimal cut sets: {1,2,3}, {4} and {5,2}. For each cut set we calculate the product of the q_i's, and thereafter sum these products:

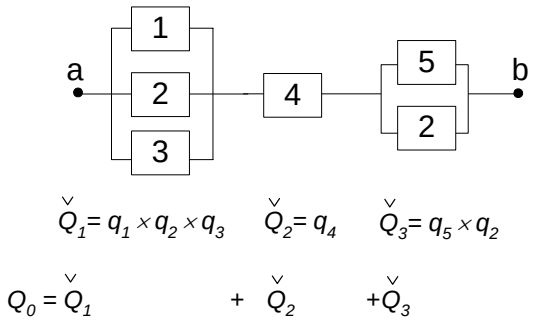


Figure 72 Calculation of Q_0 based on the minimal cut sets

A slightly better approximation is given by the ”upper bound approximation”:

$$Q_0 \approx 1 - \prod_{j=1}^{k} (1 - \breve{Q}_j) \tag{121}$$

17.7.2F_0 – TOP event frequency

We will demonstrate the method for calculating the TOP event frequency, F_0 in the following situation for each minimal cut set:

- One and only one basic event is of the type ”frequency” with occurrence rate *f.*
- The remaining basic events is of the type “barrier/on demand probability”, with a barrier probability *q.*

F_0 could now be approximated by:

$$F_0 = \sum_{\text{all cut sets } K_j} \left\{ f_{k_j} \cdot \prod_{\substack{i \in K_j \\ i \neq k_j}} q_i \right\} \tag{122}$$

where f_{kj} is the rate of the basic event of type ” frequency” in cut set K_j, and q_i is the probability that basic event i in cut set K_j occurs. Figure 73 shows a graphical illustration of the procedure. To calculate the frequency of each cut set we multiply the frequency with the barrier probabilities, and then we sum the contributions from each minimal cut set.

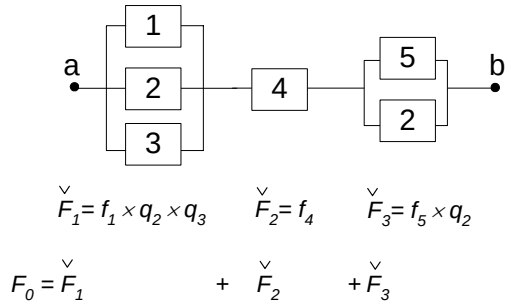


Figure 73 Calculation of F_0 based on the minimal cut sets

Note that it does not make sense to calculate F_0 in the situation where all basic events in one or more minimal cut set are of type ”probability of failure on demand” (PFD). Also note that

if there are more than one basic event of type “frequency” or “repairable” we treat one of these at a time, and calculates q for the other(s) and multiply the “frequency” with the “probabilities”, and then sum the contributions.

Exercise 25
Consider a situation were we have minimal cut sets {1,2,3} and {1,4}. Construct a fault tree that corresponds to these cut sets. Calculate F_0 when we have the following reliability parameters: $f_1 = 0.1$, $q_2 = q_3 = 0.1$, and $q_4 = 0.05$. □

Exercise 26
Consider the fault tree in exercise **24**. Assume the following parameters:

Component	λ (hrs^{-1})	MDT (hrs)
TR	$5 \cdot 10^{-6}$	720
SM	$5 \cdot 10^{-5}$	48
MDV	$4 \cdot 10^{-5}$	24
SV	10^{-3}	24
IPC	$8 \cdot 10^{-4}$	8
PLC	$5 \cdot 10^{-4}$	16
OPS	$8 \cdot 10^{-5}$	24

Find the TOP event probability Q_0, and the TOP event frequency, F_0.

17.7.3 $MTTF_0$ - Mean Time To system Failure

The following approximation could be used to calculate the mean time to system failure $MTTF_0$:

$$MTTF_0 = 1/F_0 \tag{123}$$

17.8 Measures of Importance

The reliability importance of a component in a system will generally depend on the location of the component in the system, and the reliability of the component. In the following we will describe some measures which quantify this relation. A number of different measures have been defined, and in this presentation the following measures will be described:

- Vesely-Fussell’s measure of reliability importance.
- Birnbaum’s measure of reliability importance.
- Improvement potential.
- Criticality Importance.
- Order of smallest cut set
- Birnbaum’s measure of structural importance.

17.8.1 Vesely-Fussell’s Measure of Reliability Importance

Vesely-Fussell’s measure of reliability importance for component i is defined by:

$I^{VF}(i|t_0)$ = *the conditional probability that at least one minimal cut set containing input event no. i is failed at time t_0, given that the system fails at time t_0.*

The following approximation, which is usually good, may be used to compute I^{VF}:

$$I^{VF}(i \mid t) \approx \frac{\sum_{j=1}^{m_i} \breve{Q}_j^i(t)}{Q_0} \tag{124}$$

where the upper index ^{i} means that, in the numerator, only the minimal cut sets containing input event no. *i* are considered. Then m_i is the number of minimal cut sets containing input event no. *i*.

Vesely-Fussell's measure of importance can be interpreted as the probability that the TOP event is *caused* by input event no. *i*, when it is given that the TOP event has occurred. Then by saying that "the TOP event is caused by input event no. *i*", we mean that input event no. *i* occurs and the rest of the input events in the fault tree are in such states that the TOP event occurs if and only if input event no. i occurs.

17.8.2Birnbaum's Measure of Reliability Importance

Birnbaum's measure of reliability importance for component *i* is defined as follows:

$I^B(i|t)$ = *the partial derivative of $Q_0(t)$ with respect to $q_i(t)$*

Thus an increase of $q_i(t)$ by a (small) amount, say a_i, will increase $Q_0(t)$ by an amount (approximately) a_i times $I^B(i|t)$.

In order to calculate Birnbaum's measure of reliability importance we may introduce another interpretation:

$$I^B(i|t) = \Pr(\text{"TOP-event occurs at } t_0\text{"} \mid q_i(t)=1) - \Pr(\text{"TOP-event occurs at } t\text{"} \mid q_i(t)=0) \tag{125}$$

i.e. the difference between the probabilities of the TOP-event computed under the assumptions that input event no. *i* is known to occur and is known to not occur, respectively. This difference may be interpreted as the probability that input event no. *i* is *critical* at time *t*.

17.8.3Improvement potential

The improvement potential reliability measure for component *i* is defined by:

$I^{IP}(i|t)$ = *the increase in system reliability if component i is replaced with a perfect component at time t.*

The improvement potential measure is related to Birnbaums measure by:

$$I^{IP}(i \mid t) = I^B(i \mid t) \cdot q_i(t) \tag{126}$$

17.8.4Criticality Importance

The criticality importance reliability measure for component *i* is defined by:

$I^{CR}(i|t)$ = *the probability that component i is critical for the system and is failed at time t, given that the system is failed at time t.*

The criticality importance measure is related to Birnbaums measure by:

$$I^{CR}(i \mid t) = \frac{I^{B}(i \mid t) \cdot q_i(t)}{Q_0(t)} \tag{127}$$

17.8.5Order of smallest cut set

The order of smallest cut set importance measure is defined by:

$I^{O}(i)$ = *The order of the smallest cut set containing component i*

Note that this is a qualitative measure that does not depend on the component reliabilities.

17.8.6Birnbaum's Measure of Structural Importance

Birnbaum's measure of structural importance for component *i* is defined as follows:

$B_\phi(i)$ =*the relative number of system states for which component i is* ***critical*** *for the system.*

Component *i* is *critical* if the state of the system is such that the system functions if and only if component *i* functions. A more precise definition of this measure is:

$$B_\phi(i) = \frac{\eta_\phi(i)}{2^{n-1}} \tag{128}$$

where $\eta_\phi(i)$ is the total number of *critical path vectors* for component *i*. A critical path vector for component *i* is a state vector of the other components in the system such that the system functions if and only if the *i*'th component functions. The idea behind this measure is to count the relative number of different states of the system (all *other* components than *i*) which cause component *i* to be *critical* for the system.

It can be shown that if all components have $q_i(t) = 0.5$, then $B_\phi(i) = I^{B}(i)$.

17.8.7Cut set importance

The cut set importance for cut set *j* is defined by

$I^{CI}(j)$ = *the conditional probability that a minimal cut set j is failed at time t, given that the system is failed at time t.*

Cut set importance is calculated by the formula

$$\frac{\prod_{i \in K_j} q_i(t)}{Q_0(t)} \tag{129}$$

where $Q_0(t)$ is the probability that the TOP event is occurring at time t.

17.9 Maintenance optimisation example

In this section we will as an example use a simplified process model to optimise the maintenance of each component. The process model is shown in Figure 74.

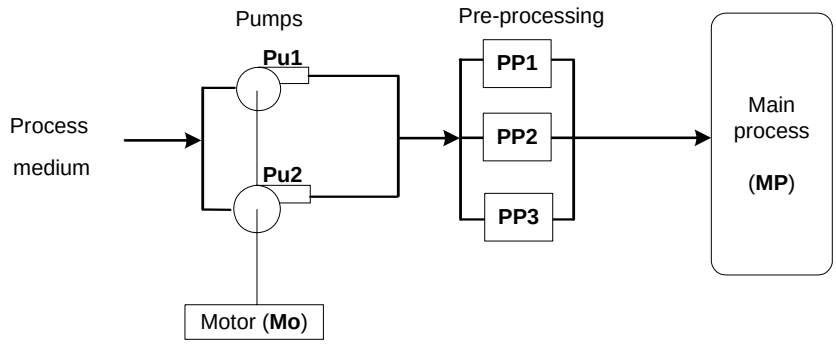


Figure 74 Simplified process model used in relation to FTA optimisation example

We make the following assumptions related to the success of the process system. The pump system requires that at least one of the two pumps function and that the motor functions. The pre-processing system comprises three pre-processing units, where at least two out of three have to function in order to support the main process. Table 33 gives relevant data for each of the components.

Table 33 Data for components in the example system

ID	MTTF	α	MDT	τ	PM_{Cost}	CM_{Cost}
Mo	15 000	3	5	8 760	30 000	35 000
PP1	2 920	1.5	24	1 460	5 000	10 000
PP2	2 920	1.5	24	1 460	5 000	10 000
PP3	2 920	1.5	24	1 460	5 000	10 000
MP	8 760	2	16	4 380	20 000	25 000
Pu1	4 370	2	48	2 185	7 000	15 000
Pu2	4 370	2	48	2 185	7 000	15 000

Where

MTTF = $MTTF_N$ = Mean Time to Failure, without preventive maintenance
α = ageing parameter
MDT = Mean Time To Repair/Mean Down Time
τ = Current maintenance interval
PM_{Cost} = Cost per preventive maintenance action (overhaul/replacement)
CM_{Cost} = Cost per corrective maintenance action (direct costs, not cost related to production loss)
U_{Cost} = 100 000 = Cost of production stop, i.e. value of lost production upon system failure

Figure 75 shows the fault tree corresponding to the model in Figure 74.

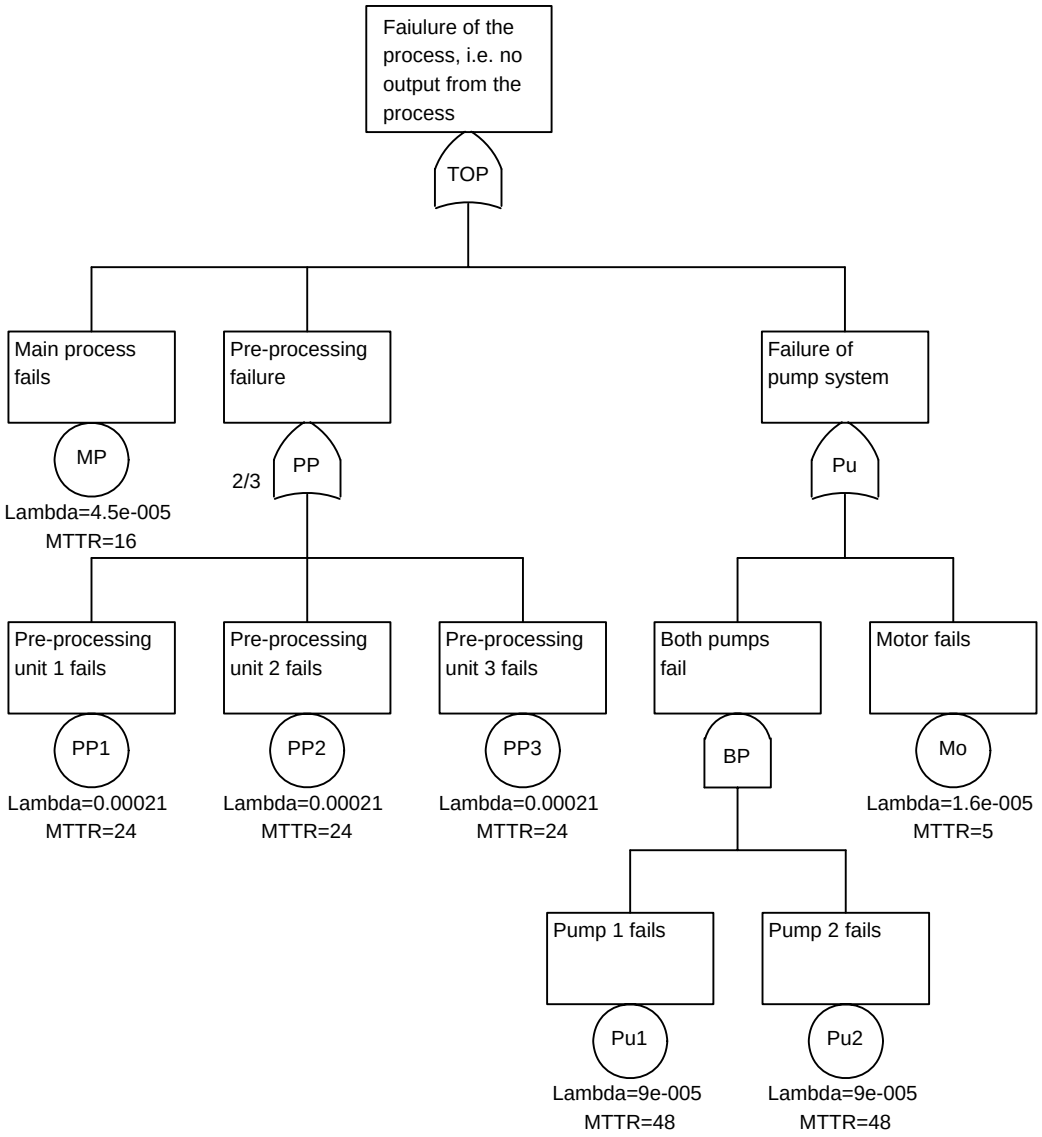


Figure 75 Fault tree for the example system

For this simple fault tree it is easy to verify that the minimal cut sets are:

Cut set(s) with 1 component (Total: 2)
{Mo}
{HP}

Cut set(s) with 2 components (Total: 4)
{PP1,PP2}
{PP1,PP3}
{PP2,PP3}
{Pu1,Pu2}

To make quantitative assessment of the fault tree we need to calculate the "effective failure rate" for each basic event (component) with the current maintenance program. The effective failure rate is approximated by:

$$\lambda_E(\tau) = \left(\frac{\Gamma(1+1/\alpha)}{\text{MTTF}} \right)^{\alpha} \tau^{\alpha-1} \qquad (130)$$

Now, using the values for MTTF, α and τ from Table 33 we obtain the "lambda" values shown on the fault tree in Figure 75. Note that the "input" parameters to the basic events are (constant) failure rates and repair times (MDT).

A computerised fault tree program easily calculates the TOP event probability, Q_0 = 8.92e-4, corresponding to 7.8 hours unavailability per year. We also obtain the frequency of the TOP event, F_0, to be 6.79e-5 [Occ. per Hours], which equals 0.6 system shut-downs per year. The Birnbaums measure of reliability importance is found to be:

Basic Event, i	Birnbaum [$I^B(i)$]
MP	9.9983e-001
Mo	9.9919e-001
PP3	9.8770e-003
PP1	9.8770e-003
PP2	9.8770e-003
Pu1	4.2977e-003
Pu2	4.2977e-003

Optimisation of maintenance intervals:
In this situation, there is three types of cost elements to include in the optimisation model; *i*) the cost of preventive maintenance which decreases with increasing maintenance intervals, *ii*) the unavailability cost which increases with increasing maintenance intervals, and *iii*) corrective maintenance cost which also increase with increasing maintenance intervals. The

cost function to optimise is the total cost per unit time:

$$C(\tau) = C_{PM}(\tau) + C_U(\tau) + C_{CM}(\tau)$$

The preventive maintenance cost per unit time is easily found by:

$$C_{PM}(\tau) = PM_{Cost}/\tau$$

Also, the corrective maintenance cost per unit time could be found by:

$$C_{CM}(\tau) = CM_{Cost} \times \lambda_E(\tau)$$

The challenge now is to obtain the unavailability cost $C_U(\tau)$. In fault tree terminology we have:

$$C_U(\tau) = U_{Cost} \times F_0$$

where F_0 is the frequency of the TOP event. F_0 is a function of the component reliability parameters, which again is a function of the maintenance interval through Equation (130). In principle, we could calculate F_0 for various values of the maintenance interval by using Equation (130) to first find the effective failure rate, and then enter these values into the Fault Tree program letting the computer calculate F_0. However, this is a tedious process which would be impractical if we have many components (basic events in the fault tree). To overcome this problem we could expand the TOP event frequency in terms of a Taylor series which yields:

$$F_0 = \sum_i I^B(i) \times \lambda_{E,i}(\tau_i)$$

where $\lambda_{E,i}(\tau_i)$ is the effective failure rate of component (basic event) i if maintained at intervals of length τ_i. Thus, when considering the maintenance of component i, we could

restrict ourselves to look at the contribution component *i* brings to the total TOP event frequency, i.e. $I^B(i) \times \lambda_{E,i}(\tau_i)$. The total cost per unite time associated to component *i* is then found to be:

$$C_i(\tau) = \frac{C_{PM_i}}{\tau_i} + \lambda_{E,i}(\tau_i)\left[I^B(i)C_U + C_{CM_i} \right] = \frac{C_{PM_i}}{\tau} + \left(\frac{\Gamma(1+1/\alpha_i)}{\text{MTTF}_\text{i}} \right)^{\alpha_i} \tau_i^{\alpha_i - 1}\left[I^B(i)C_U + C_{CM_i} \right]$$

The optimum value of τ_i is found to be:

$$\tau_i = \frac{\text{MTTF}_\text{i}}{\Gamma(1+1/\alpha_i)} \left(\frac{C_{PM_i}}{\left(I^B(i)C_U + C_{CM_i} \right)(\alpha_i - 1)} \right)^{1/\alpha_i} \tag{131}$$

Applying Equation (131) now yields a new optimised maintenance program shown in Table 34.

Table 34 Optimised vs current maintenance program for the example system

Component	Optimised maintenance program τ (hours)	Optimised maintenance program τ (months)	Current maintenance program τ (months)
Mo	8077	11.2	12
PP1	3038	4.2	2
PP2	3038	4.2	2
PP3	3038	4.2	2
MP	3954	5.5	6
Pu1	3321	4.6	3
Pu2	3321	4.6	3

18. EVENT TREE ANALYSIS

18.1 Introduction

An event tree is a logical diagram which displays possible event sequences following a specified critical event in a system. An event tree analysis (ETA) is a method for systematic analysis of a system after a critical event has occurred. The result of an ETA is a list of possible event sequences that follows the initiating event. The critical, initiating event may be a technical failure or some human error. In the development of the event sequences, the effects of possible barriers and safety functions, which are designed to prevent the occurrence of the critical event or reduce the consequences of this event, are taken into account. The analysis is both qualitative and quantitative. The qualitative content is primarily a visualisation of different scenarios (the event tree) with corresponding end consequences, while the quantitative analysis gives frequencies for the different end consequences.

18.2 Procedure

The event tree analysis is usually carried out in six steps:

1. Identification of a relevant initiating event (which may give rise to unwanted consequences).
2. Identification of the barriers and safety functions which are designed to prevent the occurrence of the initiating event, or to reduce the consequences of this event.
3. Construction of the event tree.
4. Description of the resulting event sequences.
5. Calculation of probabilities/frequencies for the identified consequences.
6. Compilation and presentation of the results from the analysis.

Each of these steps are described in the following.

18.3 Identification of Initiating Event

Selection of a relevant initiating event is very important for the analysis. The initiating event may be a technical failure or some human error. To be of interest for further analysis, the initiating event must give rise to a number of consequence sequences. If the initiating event gives rise to only one consequence sequence, fault tree analysis is a more suitable technique to analyse the problem.

The initiating event is normally identified and anticipated as a possible critical event already in the design phase. This means that barriers and safety functions have been introduced to deal with the event.

Most analysts will usually define slightly different initiating events for an analysis. For example for a safety analysis of an oxidation reactor one analyst may choose “Loss of cooling water to the reactor” as a relevant initiating event. Another analyst may choose “Rupture of cooling water pipeline” as a relevant initiating event.

18.4 Identification of Barriers and Safety Functions

The safety functions (barriers, safety systems, procedures, operator actions, etc.) that respond to the initiating event can be thought of as the system's defence against the occurrence of the initiating event. The safety functions usually include:

- Safety systems that automatically respond to the initiating event (e.g. automatic shutdown systems, automatic train protection (ATP) etc)
- Alarms that alert the operator when the initiating event occurs (e.g. fire alarm systems, alarms in the train control room)
- Operator procedures following an alarm (e.g. procedure for how to contact trains in an emergency situation)
- Barriers or containment methods that are intended to limit the effects of the initiating event (e.g. guide rails on bridges).

18.5 Construction of the event tree

The event tree displays the chronological development of states/events, beginning with the initiating event and proceeding through successes and/or failures of the safety functions that respond to the initiating event. The consequences are clearly defined outcomes of the initiating event.

The diagram starts at the left side of a page with the symbol for the initiating event. The diagram expands at each safety function, illustrated by the barrier symbol for the safety function. Within the barrier symbol the safety function is formulated as a question. To obtain a systematic diagram which is easy to read, the questions should be formulated such that the most critical output is obtained when the question is answered by "NO". The output from a barrier symbol may lead to another barrier symbol. The development is continued to the resulting consequences, illustrated by consequence symbols. If we adopt the convention that the "No" branch ("barrier fails to hold") is the downhand branch from the barrier symbol. The most severe consequences will then normally be located to bottom right corner of the consequence spectrum.

If the diagram is too big to be drawn on a single page, it is possible to isolate branches and draw them on different pages. The different pages are linked together by specific transfer symbols.

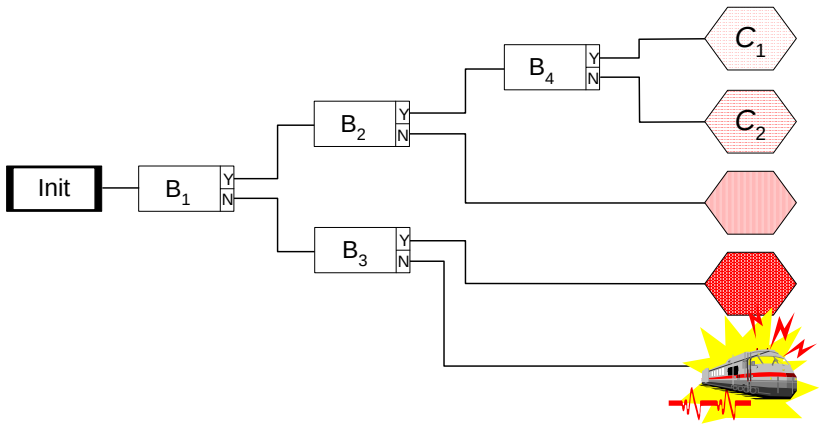


Figure 76 Example of an event tree

The “NO”-output from a barrier symbol (i.e. failure of a barrier/safety function) is often analysed by a fault tree to identify the causes for this failure. This may graphically be accomplished by linking a fault tree to the “No”-output.

An example of a very simple cause consequence diagram is shown in Figure 76.

18.6 Description of resulting event sequences

The last step in the qualitative part of the analysis is to describe the different event sequences arising from the initiating event. One or more of the sequences may represent a safe recovery and a return to normal operation or an orderly shutdown. The sequences of importance, from a safety point of view, are those that result in accidents.

The analyst must strive to describe the resulting consequences in a clear and unambiguous way. When the consequences are described the analyst may rank them according to their criticality. The structure of the diagram, clearly showing the progression of the accident, helps the analyst in specifying where additional procedures or safety systems will be most effective in protecting against these accidents.

18.7 Quantitative analysis

If relevant reliability data is available for the initiating event and all the activated safety functions, a quantitative analysis of the diagram may be carried out to give probabilities/-frequencies of the resulting consequences.

For the initiating event we usually specify the occurrence frequency of the event, i.e. the expected number of occurrences per time unit. For the various barriers/safety functions we have to specify the probability that the barrier/safety function fails to hold when activated. To assess this probability we normally have to estimate the failure rates of each of the components comprising the barrier/safety function. We also have to know how the various components are linked together and the possible maintenance strategies. The assessment may then be carried out by a fault tree analysis.

If we assume that all the barriers/safety functions are statistically independent, it is a rather simple procedure to combine the data to obtain the consequence probabilities/frequencies. These are obtained by multiplying the frequency of the initiating event by the probabilities of the relevant barrier symbols along the actual event sequence.

In order to carry out the quantitative analysis we need the frequency of the initiating event, and the barrier probabilities. During construction of the event tree, we enter the probability that the various barriers fails (i.e. the “NO” results). For each barrier, i, we enter:

$$q_i = \text{probability that barrier } i \text{ fails (”No”)} \tag{132}$$

Similarly we have:

$$p_i = 1 - q_i \text{ probability that barrier } i \text{ functions as intended (“Yes”)} \tag{133}$$

In addition to the barrier probabilities, we enter the frequency of the initiating event:

$$f = \text{frequency of initiating event} \tag{134}$$

When establishing the barrier probabilities and the initiating frequency it might be required to perform separate analyses, e.g. fault tree analysis.

To calculate the frequencies of the various consequences we may multiply the frequency of the initiating event by the barrier probabilities for each barrier along the path leading to the actual consequence[13]. Now, consider consequence C_j, and assume that S = is the set of barriers in the path leading to consequence C_j, and that represents "success" of the barrier (Yes-terminal), and further F = is the set of those barriers on the path leading to consequence C_j, and that represent "the barrier fails" (No-terminal) we have that the frequency of consequence C_j is given by:

$$F_j = f \prod_{i \in S} p_i \prod_{i \in F} q_i \tag{135}$$

To calculate equation (135) we multiply the following three factors:

- The frequency, f of the initiating event;
- $\prod p_i$ = the product of success probabilities for barriers with a "Yes" terminal (along the path from the initiating event to consequence C_j)
- $\prod q_i$ = the product of failure probabilities for barriers with a "No" terminal (along the path from the initiating event to consequence C_j)

Note that equation (135) only is valid if the barriers are independent (see discussion in footnote 13).

Exercise 27
Find the frequencies of the consequences $C_1, C_2, \ldots, C_5$ in Figure 76 when we have the following data: $f = 10$, $q_1 = 0.1$, $q_2 = 0.2$, $q_3 = 0.3$ and $q_4 = 0.1$. □

18.8 Application to railway related problems

The idea of the ETA technique is to model the situation where safety barriers are built into a system to prevent unwanted (initiating) events from developing into unwanted consequences.

Examples of initiating events in railway related problems may be:

- Signalling error (red light instead of green)
- Operator fails to recognise red light
- Level crossing failure

Examples of barriers are:

- Automatic Train Stop system
- Switch off current by control room operator
- Use of horn to warn people crossing the line in case of a level crossing failure

18.9 Result presentation

In the presentation of results we typically include:

- Listing of all identified consequences.
- Ranking of the various consequences.

[13] Note that the barrier probabilities should be specified conditionally on the outcome of previous barriers. If we are able to do this, we have taken any common cause effects into account. In practice, it is however difficult to give the conditional probabilities, and it might then be necessary to conduct a separate common cause analysis.

- Description of sequence of events for the most severe consequences.
- The occurrence frequency of each consequence.
- Evaluation of any dependencies between the barriers.
- Suggestions for improvement in terms of additional safety functions, or strengthening of weak barriers.

18.10Measure of criticality importance

In many situations we will like to establish criticality measures for each of the safety functions and barriers in the event tree. The following questions will be of importance when defining such a measure:

- How should we give weights to each of the consequences?
- How should we treat ”change” in the barrier probabilities?

Assume that it is possible to specify the “importance” of each end consequence C_j with the weight w_j. In the event tree we could then define the total “loss” related to the initiating event with

$$L = \sum_{j=1}^{n} w_j F_j \tag{136}$$

where F_j is the frequency of end consequence C_j given by equation (135). Note that the F_j’s will depend on the frequency of the initiating event, and the different barrier probabilities, q_i. We will now use an analogy to Birnbaum’s measure for reliability importance in a fault tree, and we introduce the following measure for criticality importance:

$$I^B(i) = \frac{\partial L}{\partial q_i} = \sum_{i=1}^{n} w_j \frac{\partial F_j}{\partial q_i} \tag{137}$$

In order to evaluate equation (137) we need F_i from equation (135), and we get:

$$\frac{\partial F_j}{\partial q_i} = \begin{cases} F_j / q_i \text{ if } i \in F \\ -F_j / p_i \text{ if } i \in S \end{cases} \tag{138}$$

It could be shown that:

$$I^B(i) = L(q_i = 1) - L(q_i = 0) \tag{139}$$

where $L(q_i = x)$ is the value of L in equation (136) if we set $q_i = x$.

Exercise 28

Verify equation (139). □

Exercise 29

Show that equation (139) could be adjusted to treat the “initiating event”, i.e. we could write:

$$I^B(\text{Initiating event}) = L(f = 1) - L(f = 0) \tag{140}$$

□

Exercise 30

Show that if there is an increase of q_i by a (small) amount, say a_i, the total loss, L in equation (136), will increase by an amount $a_i \times I^B(i|t)$. □

Exercise 31
Consider the event tree in Figure 77. The first barrier is the emergency shut-down. The next barrier is prevention of ignition, which depends very much on whether we are able to stop all equipment and activities that could cause an ignition, e.g. hot work, electric equipment etc. The third barrier (minor explosion) is not a real "barrier", but rather a description of the magnitude of the explosion, whereas the forth barrier (no escalation) relates to the strength of firewalls, construction etc.

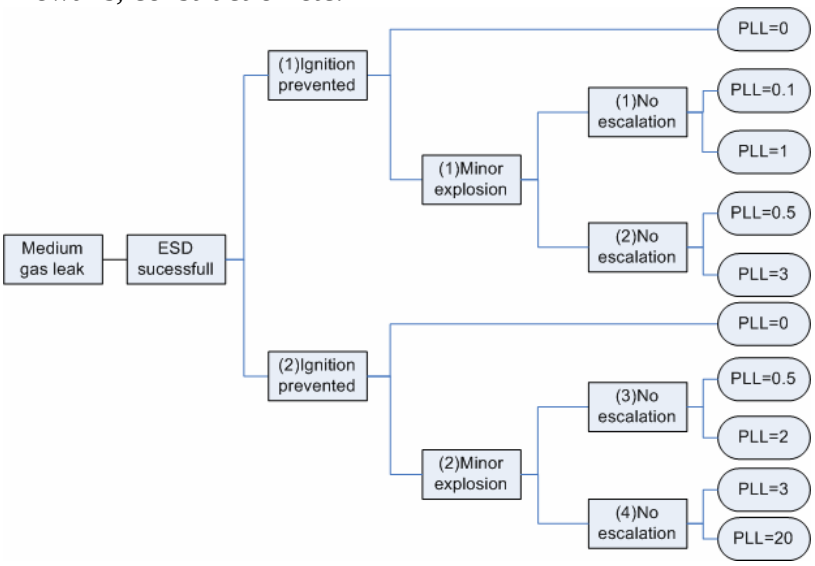


Figure 77 Event tree for gas leak situation

We will assume the following parameters

Parameter	Value
f_{INIT}	1.67
q_{ESD}	0.05
q_{IP1}	0.01
q_{IP2}	0.1
q_{ME1}	0.1
q_{ME2}	0.5
q_{NE1}	0.1
q_{NE2}	0.5
q_{NE3}	0.5
q_{NE4}	0.8

Use MS Excel to calculate the frequency, F_j, of the end consequences, and find the PLL contribution from medium gas leaks. □

Exercise 32
Consider exercise **31**, and find $I^B(i)$ for the various barriers, and the initiating event. □

Exercise 33
Consider exercise **31**, and find the increase in the PLL related to medium gas leaks if the probability of an ESD failure increases with 0.01 by *a*) using the Birnbaums measure, and *b*) using the Excel model directly. □

Exercise 34

Consider exercise **31** and assume that q_{ESD} is influenced by the quality of maintenance management system (RIF1) and the skill of maintenance staff (RIF2). A reasonable model that determines the relation between q_{ESD} and the RIFs is: $q_{ESD} = q_{ESD,0} \times w_1^{(-RIF1)} \times w_2^{(-RIF2)}$ where $q_{ESD,0} = 0.05$, $w_1 = 2$ and $w_2 = 3$. Give interpretations of the parameters $q_{ESD,0}$, w_1 and w_2. Find the PLL if RIF1 is changed from the baseline value 0 to 0.5, and RIF2 is changed from 0 to 0.8.

□

Exercise 35

We will again consider exercise **31**, but we will now make a detailed model of the ESD barrier. We will assume that the ESD function is implemented by a safety instrumented system (SIS) comprised of a 2oo3 detector system, a voting logic (CPU) and an ESD valve. The ESD valve could also be activated by an operator, e.g. if operating people smell gas, they could inform the control room in order to shut-down the process. Construct a fault tree with the TOP-event "ESD failure" in this situation. We assume the following reliability parameters:

Parameter	Value
$\lambda_{Detector}$	1e-4
$q_{Operator}$	0.60
λ_{Valve}	2e-4
λ_{CPU}	1e-5
β	0.10
PSF	0.05

The common cause parameter (β) and the probability of systematic failure (PSF) is related to the detector system. Determine the test interval such that $q_{ESD} = 0.05$ (i.e. as given in the event tree). Use the standard β factor model, but include systematic failures (PSF).

. □

Exercise 36

We will establish a "Birnbaum" like measure for a basic event of a fault tree, where the TOP event of this fault tree is a "barrier" in the event tree. Let i be the basic event in the fault tree of interest, and let j be the barrier in the event tree for which the fault tree is used to model the barrier failure probability (q). Show that $I^B(i) = I^B_{FTA}(i)\ I^B_{ETA}(j)$, where $I^B_{FTA}(i)$ denotes Birnbaums measure in the fault tree, and $I^B_{ETA}(j)$ denotes Birnbaums measure in the event tree. Discuss the result if basic event i is involved in more than one fault tree in the event tree.

□

19. MARKOV ANALYSIS

19.1 Introduction

Markov analysis is used to model systems which have many different states. These states range from “perfect function” to a total fault state. The migration between the different states may often be described by a so-called Markov-model. The possible transitions between the states may further be described by a Markov-diagram, or a state diagram.

19.2 Purpose

Markov analysis is well suited for deciding reliability characteristics of a system. Especially the method is well suited for small systems with complicated maintenance strategies. In a Markov analysis the following topics will be of interest:

- Estimating the average time the system is in each state. These numbers might further form a basis for economic considerations.
- Estimating how many times the system in average “visits” the various states. This information might further be used to estimate the need for spare parts, and maintenance personnel.
- Estimate the mean time until the system enters one specific state, for example a critical state.

19.3 Procedure

The Markov Analysis is usually carried out in six steps:

1. Make a sketch of the system
2. Define the system states
3. Group similar sates to one state (reduce dimension)
4. Draw the Markov diagram with the transition rates
5. Quantitative assessment
6. Compilation and presentation of the result from the analysis

19.4 Make a sketch of the system

The sketch is mainly used to visualise parallel and serial structures, stand-by systems, switching systems etc. In Figure 78 we have drawn a sketch of a simple cold standby system. Normally the unit A operates. If component A fails, the switch (S) activates component B.

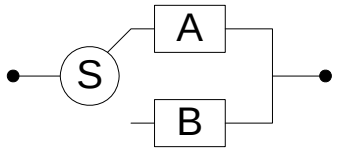


Figure 78 Example of cold standby system with switch unit S

19.5 Define the system states

Based on the sketch of the systems the various components are identified. For each component one or more states are defined. Often a number is given to each state, where the highest number represents perfect performance, whereas zero represent a complete fault state.

Next the various states of all components are combined. Each unique combination represent one system state. It is easy to imagine that this may lead to an enormous number of system states. Note that there might be combinations of component states that are not possible due to physical reasons. In Table 35 we have shown the system sates for the cold standby system together with the component states. The combination of component states depends on maintenance strategies, how the switch fails etc.

Table 35 Example of system sates for the cold standby system

System state x_S	Component state			Comments
	x_A	x_B	x_S	
4	1	1	1	All components OK
3	0	1	1	A in a fault state, the others OK
2	1	1	0	The switch is in a fault state, the others OK
1	0	1	0	A and the switch
0	0	0	1	Only the switch is OK

19.6 Group similar sates to one state (reduce dimension)

This step is only introduced in order to reduce the dimension of the problem. But in many situations several components may be identical. In this situation it will usually be possible to group similar system states into one system state, and hence reduce the dimension of the problem.

19.7 Draw the Markov diagram with the transition rates

The various system states are now drawn in a Markov diagram. Each state is drawn as a circle labelled with the state number. Transitions between the states are now visualised by drawing arrows between the corresponding circles. On each arrow the transition rate is labelled. Very often the Greek letter λ represents component failure rates, whereas the Greek letter μ represents repair rates.

19.8 Quantitative assessment

The first step is to construct the transition matrix from the Markov diagram. The transition matrix will form the basis for all quantitative assessments. The following results will often be of interest:

- The dependent solution, i.e. the state probabilities as a function of time. The state probability $P_i(t)$ is the probability that the system is in state i at time t.
- The asymptotic solution, i.e. the state probabilities when the system has reached a steady state. The steady state probability P_i represents the probability that the system is in state i when the effect of the initial state is levelled out. If state 0 represent a fault state, then P_0 represent the average unavailability of the system.
- Visiting frequencies. The visiting frequencies represent the average number of times each system state is visited per unit time. For the state defined as a "fault state", the corresponding visiting frequency represents the system failure rate.

In the following we will present the basic elements of quantitative Markov analysis. The results are given without proofs, and the reader is referred to a standard text book for a detailed description, e.g. Rausand and Høyland (2003).

19.8.1Motivating example

As a motivating example we will consider a system comprising an active pump and a spare pump in cold standby. If the active pump fails, the stand by pump is started and continue to do the duty. The failed active pump is then repaired. If the standby pump, which now is working, fails during the repair of the failed pump, we will have a system failure. The situation is illustrated in Figure 79.

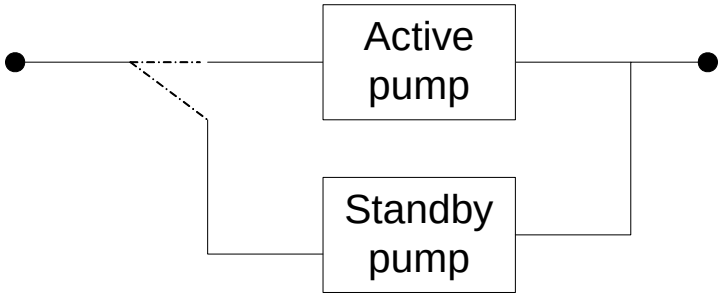


Figure 79 Pump system comprising an active pump, and a pump in cold stand by

In order to analyse this system we define the various states for each component, and for the system.

For the components (pumps) we use the following notation for component *i*

$$x_i = \begin{cases} 1 \text{ if component } i \text{ is functioning} \\ 0 \text{ if component } i \text{ is in a fault state} \end{cases} \quad (3)$$

The system as such may have more than two states, and we usually start the numbering with the highest number (all components are functioning), down to 0 which represents that all components are in a fault state. In the example let x_1 denote the state of component 1 (active pump), x_2 the state of component 2 (standby pump), and x_S denotes the state of the system. Then the following combination of states seems relevant:

Table 36 Possible states for the pump system

System state x_S	Component state		Comments
	x_1	x_2	
2	1	1	Both pumps functioning
1	0	1	The active pump is in a fault state, the standby pump is functioning
0	0	0	Both pumps in a fault state

For this system we have assumed that if the active pump fails, the standby pump could allways be started, Further we assume that if both pumps have failed, they will both be repaired before the system is put into service again.

The transition between the different system sates are now described by failure and repair rates. Introduce:

λ_1 = failure rate of the active pump

λ_2 = failure rate of the standby pump (while running, $\lambda_2 = 0$ in standby position)

μ_1 = repair rate of the active pump ($1/\mu_1$ = Mean Time To Repair when the active pump has failed)

μ_B = repair rate when both pumps are in a fault state. I.e. we assume that if the active pump has failed, and a repair with repair rate μ_1 is started, one will "start over again" with repair rate μ_B, if the standby pump also fails, independent of "how much" have been repaired on the active pump.

19.8.2Markov-diagram

The Markov (state space) diagram is shown Figure 80:

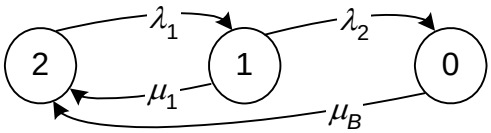


Figure 80 Markov diagram for the pump system

The circles represent the system states, and the arrows represent the transition rates between the different system states.

The Markov diagram in Figure 80 and the description in Table 36 represent the qualitative description of the system. We will now demonstrate how to perform quantitative analysis. The following quantities are of interest.

- Average time the system remain in the various system states
- The visiting frequencies to each system state
- Mean time from system start-up until the system fails for the first time

The following presentation requires knowledge of matrix algebra.

Define the transition matrix, **A**:

$$\mathbf{A} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0r} \\ a_{10} & a_{11} & \cdots & a_{1r} \\ \vdots & a_{ij} & & \\ a_{r0} & a_{r1} & \cdots & a_{rr} \end{bmatrix} \tag{141}$$

Note the following:

- The indexing starts on 0, and moves to r, e.g. it is $r+1$ system states
- Each cell in the matrix has two indexes, where the first (row index) represent the "from" state, whereas the second (column index) represent the "to" state.
- The cells represent transition rates from one state to another. a_{ij} is thus the transition rate from state i to state j.
- The diagonal elements are a kind of "dummy"-elements, which are filled in at the end, and shall fulfil the condition that all cells in a row adds up to zero.
- Since the row sum is zero, the diagonal elements a_{jj}'s are positive numbers.

The example now yields the following transition matrix: (From $\rightarrow$, To $\downarrow$):

$$\mathbf{A} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} -\mu_B & 0 & \mu_B \\ \lambda_2 & -\lambda_2 - \mu_1 & \mu_1 \\ 0 & \lambda_1 & -\lambda_1 \end{bmatrix} \end{array} \tag{142}$$

Exercise 37
Consider the pump system in Figure 79. Assume that when the active pump fails there is a constant probability p = 0.1 that the standby pump do not start. Draw the Markov diagram, and set up the transition matrix. Write down the assumptions you make. □

19.8.3Steady state probabilities

Let the vector $[P_0, P_1,\ldots,P_r]$ represent the average time the system is in the various system states. I.e. P_0 is average fraction of the time the system is in state 0, P_1 is average fraction of the time the system is in state 1 etc. The elements $\mathbf{P} = [P_0, P_1,\ldots,P_r]$ are also denoted steady state probabilities to indicate that in the stationary situation P_i represents the probability that the system is in state i. It can be shown that **P** is the solution of the following matrix equation:

$$\mathbf{P}\,\mathbf{A}_1 = \mathbf{b} \tag{143}$$

Where

$$\mathbf{A}_1 = \begin{bmatrix} a_{00} & a_{01} & \cdots & 1 \\ a_{10} & a_{11} & \cdots & 1 \\ \vdots & & & \\ a_{r0} & a_{r1} & \cdots & 1 \end{bmatrix} \tag{144}$$

is the ordinary transition matrix, but with the rightmost column replaced with only ones, and $\mathbf{b} = [0,0, \ldots,0,1]$ is a vector of r zeros and a number one on the last position.

We would prefer to solve (143) analytically, such that the P_i's may be written as an explicit function of the transition rates. In practice we often solve equation (143) with numerical methods, e.g. by Microsoft Excel.

In the example we have:

$$\begin{bmatrix} P_0 & P_1 & P_2 \end{bmatrix} \begin{bmatrix} -\mu_B & 0 & 1 \\ \lambda_2 & -\lambda_2 - \mu_1 & 1 \\ 0 & \lambda_1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \tag{145}$$

which yields

$$P_0 = \frac{\lambda_1 \lambda_2}{(\lambda_2 + \mu_B)\lambda_1 + (\lambda_2 + \mu_1)\mu_B}$$

$$P_1 = \frac{\mu_B \lambda_1}{(\lambda_2 + \mu_B)\lambda_1 + (\lambda_2 + \mu_1)\mu_B}$$

$$P_2 = \frac{\mu_B(\lambda_2 + \mu_1)}{(\lambda_2 + \mu_B)\lambda_1 + (\lambda_2 + \mu_1)\mu_B} \tag{146}$$

19.8.4Visiting frequencies

Often we are interested in evaluating how many times the system enters the various states, i.e. the visiting frequencies. The visiting frequency for state *j* is denoted ν_j, and could be obtained by:

$$\nu_j = -P_j \cdot a_{jj} \tag{147}$$

From our example we obtain e.g.:

$$\nu_0 = -P_0 a_{00} = \frac{\mu_B \lambda_1 \lambda_2}{(\lambda_2 + \mu_B)\lambda_1 + (\lambda_2 + \mu_1)\mu_B} \tag{148}$$

which could be interpreted as the “system failure rate”.

19.8.5Mean time to a given system state

In order to obtain mean time until a given system state is reached (from system start-up) we introduce the Laplace transform, see e.g. Rausand and Høyland (2004) for a detailed explanation:

1. Let *k* represent the state we will investigate. E.g. *k* = 0 corresponds to a total pump system failure in the example.
2. Take matrix **A** in equation (141) as a starting point, and delete column *k*, and row *k*. Denote the new matrix with $\mathbf{A}^r$ (*r* for reduced).
3. Solve $\mathbf{P}^*$ in the matrix equation $\cdot\mathbf{P}^* \mathbf{A}^r = \mathbf{b}$, where **b** = [0,0, …,0,-1,0, …0] is a vector of only zeros, except for position *l*, where we insert -1, and *l* is the initial system state. Note that the number of elements in **b** is also reduced from *r*+1 to *r*. This will influence the positioning of”-1”, which have to be moved ”one to the left”, if $k < l$.
4. Mean time until reaching the actual system state for the first time could now be found by summing the elements of $\mathbf{P}^*$, i.e.. $\text{MTTF}_S = \sum_i P_i^*$

Let’s assume that we are interested in obtaining mean time to total failure of the pump system, i.e. state 0. We then delete column *k* = 0, and row *k* = 0. Further let **b** = [0, -1], which yields:

$$\begin{bmatrix} P_1^* & P_2^* \end{bmatrix} \begin{bmatrix} -\lambda_2 - \mu_1 & \mu_1 \\ \lambda_1 & -\lambda_1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \end{bmatrix} \tag{149}$$

The solution is:

$P_1^* = 1/\lambda_2$

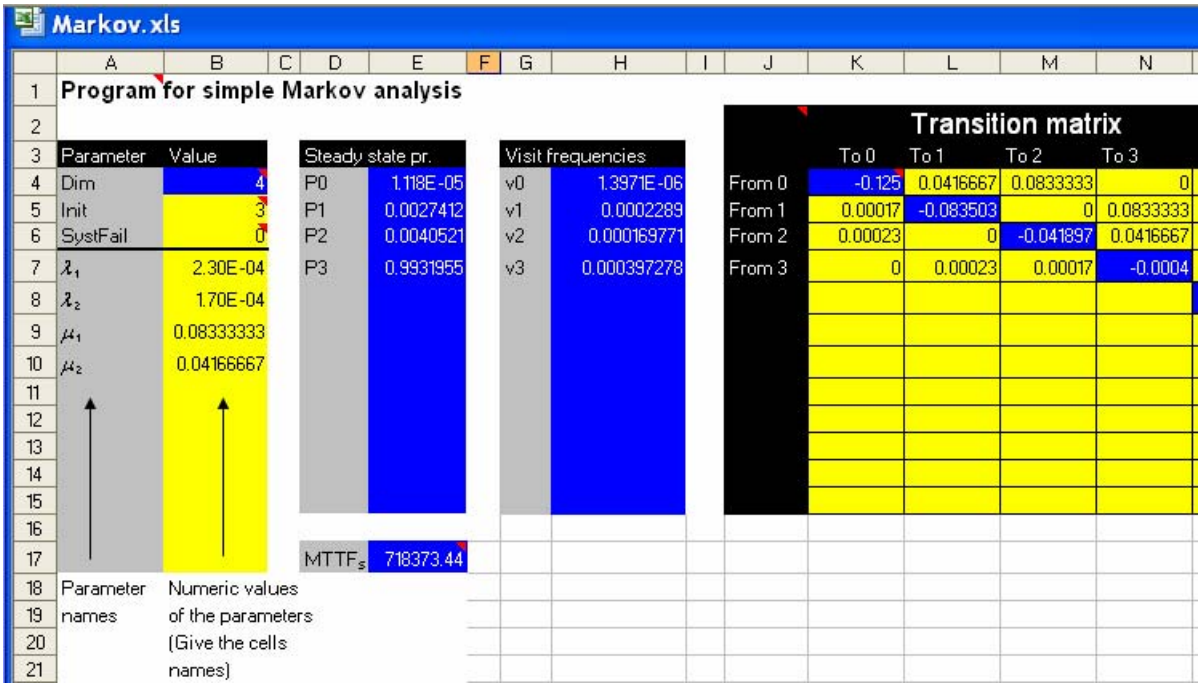
$P_2^* = (\lambda_2 - \mu_1)/(\lambda_1\lambda_2)$ (150)

Thus

$\text{MTTF}_S = P_1^* + P_2^* = (\lambda_1 + \lambda_2 + \mu_1)/(\lambda_1\lambda_2)$ (151)

Note that if $\mu_1 = 0$ (i.e. no repairs),we will have $MTTF_s = (\lambda_1+\lambda_2)/(\lambda_1\lambda_2) = 1/\lambda_1 + 1/\lambda_2 = MTTF_1 + MTTF_2$ as expected.

A simple program in Microsoft Excel has been provided to obtain the calculations:



Exercise 38

Consider the pump system in Figure 79. Now, assume that the standby pump ha a constant failure rate $\lambda_3 = 0.0005$ for failing in ”cold” standby, and a repair rate $\mu_2 = 0.05$ in this situation. Make a listing over the various sates for this system. Draw the Markov diagram, and set up the transition matrix. Find the steady state probabilities, and find the visiting frequencies and $MTTF_S$. Write down the assumptions you make. Hint: Use the Excel spreadsheet. □

Exercise 39

In this exercise we will consider a parallel system with two units A and B. In ordinary operation both units are working on half load, each having a constant failure rate equal to $\lambda_L = 0.0005$. If one of the units fails, the other unit will have to take the entire load, and hence the failure rate increases to $\lambda_H = 0.001$. After failure of any of the two units, a repair is started, with constant repair rate $\mu = 1/16$. If the operating unit also fails while the failed unit is being repaired we have a system failure. The repair of the first failed unit continues with repair rate μ, whereas the repair of the last failed unit is $\mu_F = 1/8$. Make a listing over the various sates for this system. Draw the Markov diagram, and set up the transition matrix. Find the steady state probabilities, and find the visiting frequencies and $MTTF_S$. Write down the assumptions you make. Hint: Use the Excel spreadsheet. □

19.9 Time dependent solution

Up to now we have investigated the steady state situation, i.e. the average portion of time the system is in the various system states. In some situations we also want to investigate the time dependent solution, i.e. the probability that the system is in e.g. state 0 at time *t.* We now let $P_i(t)$ be the probability that the system is in state *i* at time *t.*

$$\mathbf{P(t)}\ \mathbf{A}_1 = d\,\mathbf{P(t)}/d\,t \qquad (152)$$

where $\mathbf{P}(t) = [P_0(t), P_1(t),\ldots,P_r(t)]$ and $d\,\mathbf{P(t)}/d\,t$ is the time derivative of **P(t).** In order to solve equation (152) we need some advanced matrix algebra which is beyond the scope of this presentation. However, a numerical solution could be obtained by the Markov.xls program.

Exercise 40

Consider Exercise **39**. Use the Markov.xls program to investigate how long time we have to wait before we reach the steady state probabilities. □

20. ADDITIONAL EXERCISES WITH SOLUTIONS

Exercise 41
We will consider the maintenance and spare part strategy for a pump. The following parameters are of interest:

Parameter	Value	Explanation
MTTF	17520	Mean Time To Failure (in hours)
α	3	Age parameter
PM_{Cost}	1000	Cost per PM activity
CM_{Cost}	5000	Cost per CM activity
U_{Cost}	2000	Unavailability cost per hour
MDT	8	Mean Down Time (in hours)

Further, we assume the effective failure rate when preventive maintaining the pump with maintenance interval τ equals $\lambda_A(\tau) = \left(\frac{\Gamma(1+1/\alpha)}{\text{MTTF}}\right)^{\alpha} \tau^{\alpha-1}$.

a) Calculate the total cost for τ equal 6, 8 and 12 months (1 month = 720 hours) and find the maintenance interval that minimises the total cost.
b) Find an expression for the τ value that minimises the total cost by analytic calculus.
c) Assume that MDT could be reduced to 1 hour if an essential spare part is kept in the first line maintenance depot rather than in the central stock. The additional spare part cost will then be 500 per year. Determine the total cost in this situation, and compare with the result in a).

Solution
a) By direct calculation we find (cost per hour)

τ	$\lambda_A(\tau)$	$C_{PM}(\tau)=PM_{Cost}/\tau$	$C_{CM}(\tau) = \lambda_A(\tau)\times CM_{Cost}$	$C_U(\tau) = \lambda_A(\tau)\times U_{Cost}\times MDT$	$C(\tau) = C_{PM}(\tau)+C_{CM}(\tau)+C_U(\tau)$
6	2.47E-06	0.2315	0.0124	0.0395	0.2834
8	4.39E-06	0.1736	0.0220	0.0703	0.2659
12	9.88E-06	0.1157	0.0494	0.1582	0.3233

Thus, the total cost is minimised for $\tau = 8$ months = 5760 hours

b) The total cost per unit time is

$$C(\tau) = PM_{Cost}/\tau + \lambda_A(\tau)\left[CM_{Cost} + U_{Cost}\times MDT\right] = PM_{Cost}/\tau + \left(\frac{\Gamma(1+1/\alpha)}{\text{MTTF}}\right)^{\alpha}\tau^{\alpha-1}\left[CM_{Cost} + U_{Cost}\times MDT\right]$$

By setting $\frac{dC(\tau)}{d\tau} = 0$ we obtain the τ value that minimises the cost to be:

$$\tau = \sqrt[\alpha]{PM_{Cost}/(\alpha-1)\left[CM_{Cost} + U_{Cost}\times MDT\right]}\times\frac{\text{MTTF}}{\Gamma(1+1/\alpha)}$$ = 5644 hours = 7.7 months, which is in accordance with a).

c) We now find

τ	$\lambda_A(\tau)$	$C_{PM}(\tau)=PM_{Cost}/\tau$	$C_{CM}(\tau) = \lambda_A(\tau)\times CM_{Cost}$	$C_U(\tau) = \lambda_A(\tau)\times U_{Cost}\times MDT$	$C(\tau) = C_{PM}(\tau)+C_{CM}(\tau)+C_U(\tau)+500/8760$
6	2.47E-06	0.2315	0.0124	0.0049	0.3059
8	4.39E-06	0.1736	0.0220	0.0088	0.2614
12	9.88E-06	0.1157	0.0494	0.0198	0.2420

Thus, the minimum cost is obtain for τ = 12 months, and the total cost is reduced compared to the situation in a), even if we add the cost 500/8760 (per hour). Note that if the cost of the keeping the spare part in the first line maintenance depot increases to 1000 per year this strategy is not economical.

Exercise 42

Consider a rail that is exposited to external chocks that may cause a crack. By use of an ultrasonic inspection train the cracks could be detected before they develops to rail breakages. Find the optimum frequency of ultrasonic inspections. The following parameters could be assumed:

Parameter	Value	Explanation
f	0.05	Number of cracks per per km
$E_{P\text{-}F}$	4	Average P-F interval (in years)
$SD_{P\text{-}F}$	2	Standard deviation of P-F interval (years)
PM_{Cost}	500	Cost per km running the ultrasonic inspection train
$CM_{Cost,P}$	15 000	Cost of fixing a detected crack
$CM_{Cost,Ut}$	40 000	Cost of fixing a rail breakage
p_I	0.9	Probability that a crack is detected by a separate inspection
p_{DR}	0.01	Probability that a rail crack results in a derailment
DR_{Cost}	15 000 000	Cost of derailment

Hint: Calculate the costs for τ = 6, 12 and 18 months respectively.

Solution

Yearly cost for different value of τ is found to be:

τ (months)	τ (years)	$f\times Q(\tau)$	$C_S(\tau)$	$C_{PM}(\tau)$	$C_{CM}(\tau)$	$C(\tau)$
2	0.17	2.9E-06	0.9	3000	0.1	3001
4	0.33	2.17E-05	6.4	1500	0.9	1507
6	0.50	8.72E-05	25.8	1000	3.5	1029
8	0.67	0.000202	59.8	750	8.1	818
10	0.83	0.000391	115.6	600	15.6	731
12	1.00	0.000651	192.6	500	26.1	719
14	1.17	0.001006	297.7	429	40.3	767
16	1.33	0.001387	410.4	375	55.5	841
18	1.50	0.001809	535.0	333	72.4	941
20	1.67	0.002411	713.0	300	96.4	1109
22	1.83	0.003029	895.8	273	121.1	1290

Thus, the optimum interval is τ = 12 months.

Exercise 43

Consider the pump system in section 19.8.1. Assume that the failure rate, λ_1 is the effective failure rate of the active pump. Thus we will write $\lambda_1 = \lambda_A(\tau)$ where τ is the length of the interval between preventive maintenance of the active pump.

Parameter	Value	Explanation
MTTF	17 520	Mean Time To Failure (in hours) pump 1 (active pump)
α	2	Age parameter pump 1
λ_2	0.01	Failure rate of standby pump (when running)
μ_1	0.125	Repair rate, pump 1 (active pump)
μ_B	0.04	Repair rate, when both pumps are repaired
PM_{Cost}	1 000	Cost per PM activity of pump 1
$CM_{Cost,1}$	4 000	Cost per CM activity of pump 1
$CM_{Cost,B}$	6 000	Cost per CM activity, repairing both pumps
$U_{Cost(1)}$	500	Unavailability cost per hour when in state 1
$U_{Cost(0)}$	10 000	Unavailability cost per hour when in state 0

a) Use the analytical formulas for $P_0, P_1, P_2, \nu_0, \nu_1$ and ν_2 to find an expression for the yearly cost of operating the pump system, i.e the cost of preventive and corrective maintenance, and the unavailability cost.
b) Calculate the total cost for τ = 3, 6 and 9 months, and propose a maintenance interval which minimises the total costs.

Solution
a) The formula for the total cost is given by

$$C(\tau) = PM_{Cos/}\tau + \nu_1 \times CM_{Cost,1} + \nu_0 \times CM_{Cost,B} + P_1 \times U_{Cost(1)} + P_0 \times U_{Cost(0)}$$

b) The total cost is found to be

τ (months)	τ (hours)	$C(\tau)$
1	730	12436
2	1460	6873
3	2190	5309
4	2920	4745
5	3650	4581
6	4380	4618
7	5110	4768
8	5840	4990
9	6570	5260
10	7300	5563
11	8030	5890
12	8760	6235

Thus, the optimum maintenance interval is τ = 5 months.

Exercise 44
Consider the event tree in section 18.5. Assume the following reliability parameters: Assume that barrier B_1 is a safety function that is periodically tested (functional test) with test interval τ. MTTF for this function is 3 years. Other relevant reliability parameters are given by: f = 0.3 per year, q_2 = 0.2, q_3 = 0.3 and q_4 = 0.1. Assume the following cost figures: AC_1 = 100, AC_2 = 500, AC_3 = 2 000, AC_4 = 50 000, AC_5 = 250 000, and PM_{Cost} = 1 000. Here AC_i is the

accident cost corresponding to consequence C_i, and PM_{Cost} is the cost of a functional test of safety barrier B_1.

Find the optimum interval for functional test of safety barrier B_1.Hint: Calculate the costs for τ = 3, 6 and 9 months.

Solution
For barrier B_1 we have barrier $q_1 = \tau/(2\text{MTTF})$, and the total cost per year is given by:

$$C(\tau) = 12PM_{Cost}/\tau + f[(1- q_1)(1- q_2)(1- q_3)AC_1 + (1- q_1)(1- q_2)q_3 AC_2 + (1- q_1)q_2 AC_3 + q_1(1- q_3)AC_4 + q_1 q_3 AC_5]$$

The cost if found to be:

τ (months)	$C(\tau)$
3	5522
4	4978
5	4835
6	4891
7	5061
8	5303
9	5593

Hence, the optimum test interval is τ = 6 months.

Exercise 45
Consider a railway system that is "running into" the end of the global bath tube curve. This means that the variable cost, $c(t)$ of operating the railway system is increasing. We will calculate the cost-benefit ratio of a minor rehabilitation program. By executing this program, the time before the system has to be renewed could be extended from 3 to 10 years. Further the variable costs, $c(t)$ could be reduced by a factor of 3. The cost of the rehabilitation program is PC = 500 000. The yearly cost, $c(t)$ is 30 000 at the moment. The increase in yearly cost is g = 10% per year. The cost of a full renewal is RC = 2 500 000. The time interval between system renewal is as a baseline 50 years. Calculate the cost benefit ratio for the rehabilitation project if the interest rent, r is set to 4%.

Solution
First we calculate the net present value of the variable cost for the first year, the second year etc. For the situation with the project, c(First year) = 10 000, for the next year we multiply with a factor f= (1+g)/(1+r). Then this amount is multiplied with f again for the next year and so forth. If the project is not executed, c(First year) = 30 000, and we multiply with f for two more years. It is assumed that after a full renewal, the variable costs could be neglected. Also, we do not include increasing variable cost when we are approaching the next renewal in 50 years time, even if we in principle could have done this. If the project is run, we have to pay the project cost, PC. Finally, we have to calculate the cost of the portfolios of renewals. For a renewal in year T, the net present value of the renewal cost (RC) is :$RC(1+r)^{-T}$. For the situation with the project we have to calculate this amount for T = 10, T = 60 etc, and if the project is not executed, we have to calculate for T = 3, T = 53 etc. The results are summarised below:

Cost element	With project	Without project
c(First year)	10 000	30 000
:	10 577	31 731
:	11 187	33 561
:	11 833	
:	12 515	
:	13 237	
:	14 001	
:	14 809	
:	15 663	
c(10th year)	16 567	
$\sum c(t)$	130 388	95 292
Project Cost	500 000	0
RC First	1 688 910	2 222 491
RC Second	237 651	312 733
RC Third	33 440	44 005
RC Forth	4 705	6 192
RC Fifth	662	871
Total	2 595 758	2 681 584

And we see that it is good economy in running the rehabilitation project with the given data.

REFERENCES

Anderson, R. T. and L. Neri. 1990. *Reliability-Centered Maintenance. Management and Engineering Methods*. Elsevier Applied Science, London.

Ascher, H. and H. Feingold. 1984. *Repairable Systems Reliability; Modeling, Inference, Misconceptions and Their Causes*. Marcel Dekker, Inc., New York.

Aven, T. and B. Bergman, B. 1986. *Optimal replacement times, a general set-up*. J. Appl. Probab, 23, pp. 432-442.

Aven, T. 1985. Reliability/Availability Evaluations of Coherent Systems Based on Minimal Cut Sets. *Reliability Engineering*, 12:93-104.

Aven, T. 1992. *Reliability and Risk Analysis*. Elsevier Science Publishers, London.

Aven, T. and U. Jensen. 1999. *Stochastic Models in Reliability*. Springer-Verlag New York.

Barlow, R. and L.C. Hunter. 1960. *Optimum Preventive Maintenance Policies*. Operations Research, 8, pp. 90-100.

Blache, K. M. and A. B. Shrivastava. 1994. Defining failure of manufacturing machinery & equipment. *Proceedings Annual Reliability and Maintainability Symposium*, pages 69-75.

Blanchard, B. S. and W. J. Fabrycky. 1981. *System Engineering and Analysis*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632.

Bodsberg, L. 1993. *VULCAN - A vulnerability calculation method for process safety systems*. PhD thesis, Norwegian Univiersity of Science and Technology. ISBN 92-7119-581-6.

BS 4778. *Quality vocabulary*. British Standards Institution, London, 1991.

BS 5760-5. 1991. *Reliability of systems, equipments and components; Part 5: Guide to failure modes, effects and criticality analysis (FMEA and FMECA)*. British Standards Institution, London

Cho, D.I and M. Parlar. 1991. *A survey of maintenance models for multi-unit systems*. European Journal of Operational Research, 51, p. 1-23.

Cox, D. R. and H. D Miller. 1965.*The Theory of Stochastic Processes*. Methuen, London.

Cross, N. 1994. *Engineering Design Methods: Strategies for Product Design*. John Wiley & Sons, Chichester.

Dekker, R. 1992. *A general Framework for Optimisation, Priority Setting, Planning and Combining Maintenance Activities*. Technical Reprot 9270/A, Econometric Inst. Erasmus Univ., Rotterdam.

EN 292-1. 1991. *Safety of machinery - Basic concepts, general principles for design - Part 1: Basic terminology, methodology*. CEN.

EN 50126. 2000. *Railway applications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)*. CENELEC.

Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. 1996. *Advances in Knowledge Discovery and Data Mining*. AAAI Press / The MIT Press, Menlo Park, California.

Hansen, G. K. and R. Aarø. 1997. Reliability Quantification of Computer-Based Safety Systems. An Introduction to PDS. Technical Report STF38 A97434, SINTEF Industrial Management, N-7465 Trondheim, Norway.

Hansen, G. K. and J. Vatn. 1998. Reliability Data for Control and Safety Systems. 1998 Edition. Technical Report STF38 A98445, SINTEF Industrial Management, N-7465 Trondheim, Norway.

Hendrick, K and L. Benner Jr. 1987.*Investigating Accidents with STEP*. Marcel Dekker inc., New York.

Hoch, R. R. 1990. A Practical Application of Reliability Centered Maintenance. *The American Society of Mechanical Engineers, 90-JPGC/Pwr-51, Joint ASME/IEEE*

Power Gen. Conf., Boston, MA, 21-25 Oct. 1990.
Hokstad, P.R. 1998. Life Cylce Cost Anslysis in Railway Systems. Technical Report STF38 A98424, SINTEF Industrial management. ISBN 82-14-00450-0.
IEC 50(191). 1990 *International Electrotechnical Vocabulary (IEV) - Chapter 191 - Dependability and quality of service*. International Electrotechnical Commission, Geneva, 1990.
IEC 60812. *Analysis Techniques for System Reliability - Procedures for Failure Modes and Effects Analysis (FMEA)*. International Electrotechnical Commission, Geneva, 1985.
IEC 61025. 1990. *Fault Tree Analysis (FTA).* International Electrotehcnical Commisssion, Genveva.
IEC 61508. 1998. *Functional safety of electrical/electroing/programmable electronic safety-related systems*. International Electrotechnical Commission, Geneva. Draft.
ISO 14224. 1999. *Petroleum and natural gas industries - Collection and exchange of reliability and maintanance data for equipment*. International Standards Organisation.
Kaplan. S. 1991. *Risk Assessment and Risk Management - Basic Concepts and Terminology*. Hemisphere Publ. Corp., Boston, Massachusetts, USA. In Risk Management: Expanding Horizons in Nuclear Power and Other industries, pp. 11-28.
Kirwan B. and L. K. Ainsworth. 1992. *A Guide to Task Analysis*. Taylor & Francis, London.
Kirwan, B. 1992. Human error identification in human reliability assessment. Part 1: Overview of approaches. *Applied Ergonomics*, 23(5):229-318.
Malik, M.A. 1990. Reliable preventive maintenance scheduling. *AIEE Trans.*, 11:221-228.
Martz, H. F. and R. A. Waller. 1982. *Bayesian Reliability Analysis*. John Wiley & Sons, New York.
Moss, M. A. 1985. *Designing for Minimal Maintenance Expense. The Practical Application of Reliability and Maintainability*. Marcel Dekker, Inc., New York.
Moubray, J. 1991. *Reliability-centred Maintenance*. Butterworth-Heinemann, Oxford.
NASA. 2002. *Fault Tree Handbook with Aerospace Applications*. NASA Office of Safety and Mission Assurance, Washington, DC.
Nolan, D.P. 1994. *Application of HAZOP and What-if safety reviews to the petroleum, petrochemical and chemical industries*, Noyes Publications 1994
Nowlan, F. S. and H. F. Heap. 1978. Reliability-centered Maintenance. Technical Report AD/A066-579, National Technical Information Service, US Department of Commerce, Springfield, Virginia.
NS 5814. 1991. *Norwegian Standard 5814. Risk Analysis Requirements*. Norwegian association of standardisation, Po.Bo 7020, 0306 Oslo, Norway.
NUREG-0492. 1981. *Fault Tree Handbook*. U.S. Nuclear Regulatory Commission, Washington, DC.
OREDA-2002. *Offshore Reliability Data*. Distributed by Det Norske Veritas, P.O.Box 300, N-1322 Høvik, Norway, forth edition. Prepared by SINTEF Industrial Management. N-7465 Trondheim, Norway.
Paglia A.M., D.D. Barnard, and D.E. Sonnett. 1991. A Case Study of the RCM Project at V.C. Summer Nuclear Generating Station. *4th International Power Generation Exhibition and Conference, Tampa, Florida, US*, 5:1003-1013.
Pierskalla, W.P. and J.A. Voelker. 1979. *A survey of maintenance models: the control and surveillance of deteriorating systems.* Nav. Res. Log. Quat., 23, p. 427-432.
Pahl, G. and W. Beitz. 1984. *Engineering Design*. The Design Council, London.
Rasmussen, J. 1986. *Information Procesing and Human-Machine Interaction*. North-Holland, Amsterdam.
Rausand, M and A. Høyland 2003. *System Reliability Theory, Models, Statistical Models, and Applications*. John Wiley & Sons, New York.

Rausand, M. 1991. *Risikoanalyse. Veiledning til NS 5814*. Tapir Forlag, N-7465 Trondheim, Norway

Rausand, M. and J. Vatn. 1997. Reliability Centered Maintenance. In C. G. Soares, editor, *Risk and Reliability in Marine Technology*. Balkema, Holland.

Reason, J. 1990. *Human Error*. Cambridge University Press, Cambridge.

Reason, J. 1997. *Managing the Risks of Organizational Accidents*. Ashgate Publishing Limited, Hampshire, 1997

Sandtorv, H. and M. Rausand. 1991. RCM - closing the loop between design and operation reliability. *Maintenance*, 6, No.1:13-21.

Smith, A. M. 1993. *Reliability-Centered Maintenance*. McGraw-Hill, Inc, New York.

Smith, D. J. 1993. *Reliability, Maintainability and Risk, Practical methods for engineers*. Butterworth Heinemann, Oxford, 4th edition.

Spjøtvoll, E. 1985. Estimation of failure rate from reliability data bases. In *Society of Reliability Engineers. Symposium* (Trondheim).

Stevens, S.S. 1946. *On the theory of scales of measurement*, Science 161, pp. 677 - 680.

Trager, Jr. T. A. 1985. *Case Study Report on Loss of Safety System Function Events*. AEOD/C504, U.S. Nuclear Regulatory Commission, Washington, DC.

Valdez-Flores, C. and R.M. Feldman. 1989. A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Naval Research Logistics Quarterly*, 36:419-446.

Vatn, G. Å., R. Rosness, and T. Paulsen. 1997. Prosedyreutvikling. Metode for analyse og beskrivelse av arbeidsoppgaver. Technical Report STF38 A97411, SINTEF Industrial Management, N-7465 Trondheim, Norway.

Vatn, J 1993. OREDA Data Analysis Guidelines. Technical Report STF38 A93024, SINTEF Industrial Management, N-7465 Trondheim, Norway..

Vatn, J. 1995. Maintenance Optimization from a Decision Theoretical Point of View. In *Proceedings, ESREL'95*, pages 273-285, London, 1995. Chameleon Press Limited.

Vatn, J. 1996. *Heterogeneity of Weibull Samples.* ESREL 1996. Crete.

Vatn, J. 1998. A discussion of the acceptable risk problem. *Reliability Engineering and System Safety*, 61(1-2):11-19, 1998.

Vatn, J. P. Hokstad, and L. Bodsberg. 1996. An overall model for maintenance optimization. *Reliability Engineering and System Safety*, 51:241-257.

Wintle, J. B, B.W Kenzie, G.J Amplett and S. Smalley. 2001. Best practice for risk based inspection as a part of plant integrity management. HSE. ISBN 0 7176 2090 5.

Øien, K. and P. Hokstad. 1998. Handbook for performing expert judgment. Technical Report STF38 A98419, SINTEF Industrial Management, N-7465 Trondheim, Norway.

Øien, K. and. R. Rosness. 1998. Methods for Safety Anlysis in Railway Systems. Technical Report STF38 A98426, SINTEF Industrial management. ISBN 82-14-00452-7.

Wang, H. 2002. *A survey of maintenance policies of deteriorating systems.* European Journal of Operational Research, 139, pp. 469-489.

Wildeman, R.E. 1996. *The art of grouping maintenance*. PhD thesis. Erasmus University Roterdam. Tindbergen Institute Research Series.

APPENDIX A – CALCULATION OF $Q_{PF}()$

In this Appendix we will describe the method used for calculating the probability that a potential failure is not detected by the inspection regime. There are two main sources for not detecting a potential failure in due time; *i*) the inspection interval is to long compared to the PF interval, and *ii*) the quality of the inspection is to low to detect a potential failure. The following quantities are defined:

T_{PF}	PF interval (random variable).
ξ_{PF}	Probability distribution function of T_{PF}
q	Failure probability of *one* inspection
q_C	Common cause part of q
q_I	Independent part of q
τ	Inspection interval

The probability that the inspection strategy fails to reveal a potential failure before a critical failure occurs could be found by the low of total probability:

$$Q_0(\tau, q, \xi_{PF}) = \int_0^\infty Q_t(\tau, q, t)\xi_{PF}(t)\mathrm{d}t \tag{153}$$

where $Q_t(\tau,q,t)$ is the probability of not detecting a potential failure given that the PF interval, T_{PF}, equals t. In order to calculate $Q_t(\tau,q,t)$ we observe that when $T_{PF} = t$, then number of possibilities to detect a failure equals n or $n+1$ where $n = \mathrm{int}(t/\tau)$ and int(·) is the integer function. The probability that we will have $n+1$ possibilities equals $t/\tau - n$ and thus the probability that we will have n possibilities to detect a potential failure equals $n + 1 - t/\tau$. Since the probability that a given inspection fails to detect a potential failure equals q, $Q_t(\tau,q,t)$ could easily be obtained by:

$$Q_t(\tau,q,t) = (n + 1 - t/\tau)\times q^n + (t/\tau - n)\times q^{(n+1)} \tag{154}$$

if the inspections could be considered statistically independent. However, the assumption that inspections are independent does not seem realistic. A more realistic assumption would be to assume that the failure probability of one inspection is given by:

$$q = q_C + q_I \tag{155}$$

where q_C represents common cause failures due to systematic failures such as low coverage, and q_I represents the failure probability due to specific conditions for one run, e.g. inadequate velocity of the measuring wagon, human errors etc.

Assuming that the failure probability of one inspection could be divided into a common and an independent part as shown in Equation (155) we calculate the total failure probability of the inspection strategy as:

$$Q_0'(\tau,q_C,q_I,\xi_{PF}) = 1-(1-q_C)(1-Q_0(\tau,q_I,\xi_{PF})) \tag{156}$$

where $Q_0(\tau,q_I,\xi_{PF})$ is found by Equation (153).

$Q_0(\tau,q_I,\xi_{PF})$ could easily be approximated by an EXCEL spreadsheet function. In the present study ξ_{PF} is assumed to be a gamma distribution with mean μ and standard deviation σ.