Maintenance Models for Systems subject to Measurable Deterioration

ISBN: 978 90 5170 997 1

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **420** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Maintenance Models for Systems subject to Measurable Deterioration

Onderhoudsmodellen voor systemen onderhevig aan meetbare veroudering

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 27 maart 2008 om 11:00 uur

door

Robin Pieter Nicolai
geboren te Rotterdam.

Promotiecommissie

Promotor:	Prof.dr.ir. R. Dekker
Overige leden:	Dr. J.B.G. Frenk
	Prof.dr. P.J.F. Groenen
	Prof.dr. M.J. Newby

Preface

After I received my MSc degree in Econometrics, I was not sure what to do next. Would I look for a job in business or in science? It took me some time to make a decision, but suddenly I was offered the opportunity to work on the DYNAFORM project at Erasmus University Rotterdam. Although I was not very familiar with the topic of this project (maintenance), I decided to give it a shot. Fortunately, I liked the project and after some time I decided to pursue a PhD degree and write a thesis on maintenance models. Now looking back at my time at Tinbergen Institute and Econometric Institute, I can say it was the right decision. The academic world appeared to be a very stimulating working environment, where I felt very comfortable. Both inside as well as outside this environment several people contributed to this thesis and below I would like to address some words to them.

First of all I would like to thank my supervisor Rommert Dekker. He was involved in almost all studies included in this book. During the last few years, Rommert always took time to read my work and answer my questions. Moreover, he always shared his research ideas and his practical experience with me. The latter helped me to approach problems from a "practical point of view". I am very happy that he stimulated me to present my work at international conferences, because conference visits in turn motivated me to start new research projects. Finally, I thank Rommert for giving me the freedom to work with other people as well.

I would like to thank Hans Frenk for co-authoring Chapters 5 and 6. He taught me how to approach problems from a "probabilistic point of view". I admire his enthusiasm for solving seemingly unsolvable problems. The last two years Hans has put a lot of time and energy in our research projects and every time I thought a problem was too difficult (or solved!), he came up with new ideas. Another co-author who deserves special attention is Alex Koning. He has taught me the concept of "structural thinking", which helped me to understand better point processes and filtrations. Sometimes our discussions were hilarious, but at the end we both exactly knew what to do. Alex, thank you for your patience and for your enormous effort to improve Chapter 3.

I thank Jan van Noortwijk for co-authoring Chapter 2 and for inviting me for his course "Sustainability Modelling" in Delft. Both his course as well as his work on maintenance

models, in particular on gamma processes, have been a source of inspiration to me. Jan, I am looking forward to working with you at HKV.

I also thank Patrick Groenen and Martin Newby, members of the small committee, for evaluating an earlier draft of this thesis. Martin, I enjoyed our (Dutch e-mail) conversations on Volterra equations and I hope we can write a paper on the evaluation of condition-based maintenance policies soon. Thanks go out as well to Albert Wagelmans and Christophe Bérenguer for being members of the big committee.

Next, I thank Nanda Piersma for encouraging me to do academic research. I would like to thank Carolina Meier-Hirmer, Gabriella Balke-Budai, Mark Vreijling and Wim Bonestroo for their useful comments and suggestions on earlier versions of Chapter 2. I thank Jan Brinkhuis who advised me to use a backward algorithm for the numerical computation of approximation 2 in Chapter 5. I gratefully acknowledge financial support of Vereniging Trustfonds EUR, Econometric Institute and Tinbergen Institute, which enabled me to present my work at conferences abroad. Many thanks also go to HKV for giving me both a warm welcome as well as the opportunity to continue working on maintenance optimization.

One of the main reasons why university life was pleasant is that I had great colleagues (even outside Erasmus University). Thank you Bram, Daniel, Dennis, Joost, Ludo, Merel, Martijn and Remco for several interesting discussions during coffee breaks, lunch or dinner. Thank you Mariëlle for many pleasant chats, self-made tiramisu and a graphical proof of Lemma 6.11. Special thanks go to Gabriella for being a true friend and being optimistic under all circumstances. Carolina also deserves another thank you; I have very much enjoyed our e-mail correspondence on various topics. Thank you Michiel for being a great friend and always listening very carefully to me.

Two very special colleagues are Rene and Wilco. They answered all my LaTeX questions and helped me to create fancy figures (even in weekends!). Rene, I never said it, but nobody could wish for a better office mate! Although Wilco was not an office mate, he could often be found in my office. At times we had great fun, but we also worked on serious OR problems. Moreover, I was very fortunate to be the first to hear his latest research ideas, even when he was in Florida. Wilco, thanks for everything!

Some people outside the university helped me forget research problems. I thank Bart de Theije for pleasant times in Rotterdam, Schiedam en Leiden (often together with Michiel). I enjoyed having dinners and watching football matches with Maarten Amelink. I would like to thank Roger Warmenhoven for being a great advisor. You did not only make me a better tennis player, but you also taught me important life lessons!

Finally, I would like to thank my mother for her never-ending support and her belief in me.

RPN

Contents

Nomenclature

For clarity we first present the terminology to be used in this thesis. We distinguish between concepts, abbreviations and mathematical notation.

Concepts

Some of the words and concepts in this thesis can be interpreted in several ways. Below we define how we use these concepts here.

Black-box deterioration model – Deterioration model that describes the relation between time and failure (e.g. hazard rate or lifetime distribution model).

Brownian motion with a time transformation – A Brownian motion with a time transformation is obtained from Brownian motion via a transformation of the clock time.

Compensator – Cumulative intensity rate.

Deterioration – The falling from a higher to a lower level in the condition of a system. More deterioration implies a worse condition level.

Discrete event system – Dynamic system that results in the occurrence of events at particular points in time, where each event has a certain discrete type.

Economic dependence – Economic dependence implies that grouping maintenance actions either saves costs or results in higher costs, as compared to individual maintenance.

Grey-box deterioration model – A grey-box deterioration model is based on a measurable quantity indicating time-dependent deterioration and failure (see *stochastic process*).

Hazard / intensity / failure rate – The instantaneous probability (or risk) of the occurrence of an event (such as failure).

Imperfect maintenance / partial repair – Maintenance that does not restore the condition of the system to as new.

Implied lifetime – The lifetime implied by the stochastic deterioration model under consideration. It is often defined in terms of a (randomized) hitting time.

Kernel density estimation – Method for the estimation of the probability density function of a random variable from a sample of data.

Maintenance optimization – The process that attempts to find those maintenance times and actions such that some decision criterion (such as the expected maintenance costs) is optimized.

Marked point process – Stochastic process that consists of events of a certain type (the mark) taking place at certain points of time.

Measurable deterioration – Deterioration that can be expressed through a *continuous* measurable quantity for which intervention limits can be set.

Multi-component maintenance – Maintenance of systems consisting of multiple components.

Non-stationary gamma process – A non-stationary gamma process is a gamma process with a non-linear shape function.

Randomized hitting time – The first time a stochastic process (such as the gamma process or Brownian motion) exceeds a random threshold.

Renewal-type equation – Recursive integral equation.

Stationary process – Stochastic process with identically distributed increments.

Stochastic approximation – Method for the optimization of objective functions with stochastic noise via gradient approximation.

Stochastic dependence – Stochastic dependence occurs if the condition of components influences the lifetime distribution of other components.

Stochastic process – Probabilistic model that incorporates temporal variability.

Structural dependence – Structural dependence applies if components structurally form a part, so that maintenance of a failed component implies maintenance of working components.

White-box deterioration model – In a white-box deterioration model the physics of measurable deterioration and failure are simulated.

Abbreviations

BM	–	Brownian motion
CDF	–	cumulative distribution function
CM	–	corrective maintenance
CV	–	coefficient of variation
DPI	–	direct plug-in
FD	–	finite differences
FDSA	–	finite differences stochastic approximation
FFT method	–	fast Fourier transform method
GP	–	gamma process
LEM	–	lifetime extending maintenance
LHS	–	left hand side
LSt-able function	–	function having a Laplace-Stieltjes transform
MIP	–	mixed integer programming problem
NHPP	–	non-homogeneous Poisson process
NTD	–	network topology dependence
PDF	–	probability density function
pLSt	–	probability Laplace-Stieltjes transform
PM	–	preventive maintenance
PMP	–	periodic maintenance problem
PMSP	–	preventive maintenance scheduling problem
RHS	–	right hand side
SA	–	stochastic approximation
SP	–	simultaneous perturbation
SPSA	–	simultaneous perturbation stochastic approximation
SSD	–	sum of squared differences
TSHG process	–	two-stage hit-and-grow process
UDE	–	univariate density estimation

Mathematical notation

Random variables are denoted by Roman capitals, constants by small Roman letters and parameters by small Greek letters. Vectors are written in bold.

$\mathbb{N}$ – the set of integer numbers $1, 2, \ldots$

$\mathbb{R}$ – the set of real numbers

$\mathbb{P}(X \leq x)$ – probability that random variable X is less than or equal to x

$\mathbb{E}(X)$ – expected value of random variable X

$\mathrm{Var}(X)$ – variance of random variable X

$cov(X, Y)$ – covariance between X and Y

$CV(X) = \sqrt{\mathrm{Var}(X)}/\mathbb{E}(X)$ – coefficient of variation of random variable X

$X \stackrel{d}{=} Y$ – the random variables X and Y have the same CDF

$X \sim F$ – the random variable X has CDF F

$1_{\{X \leq x\}}$ – $\{0, 1\}$ random variable given by

$$1_{\{X \leq x\}} = \begin{cases} 1 & \text{if } X \leq x \\ 0 & \text{otherwise} \end{cases}$$

$1_{\mathcal{A}} : [0, T] \to \mathbb{R}$ – the indicator function of the set $\mathcal{A} \subseteq \mathbb{R}$ given by

$$1_{\mathcal{A}}(t) = \begin{cases} 1 & \text{if } t \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{F}(x)$ – the fractional part of x given by $\mathcal{F}(x) = x - \lfloor x \rfloor$

$v^{\leftarrow}$ – the inverse function of the function v

$t \mapsto f(t)$ – f is a function of t

$x-$, $x+$ – $\lim_{\epsilon \downarrow 0} x - \epsilon$, $\lim_{\epsilon \downarrow 0} x + \epsilon$

$\Phi(x) = \int_{u=-\infty}^{x} (2\pi)^{-1/2} \exp(-\frac{u^2}{2}) \mathrm{d}u$ – CDF of standard normal random variable

$\Gamma(a) = \int_0^\infty x^{a-1} \exp(-x) \mathrm{d}x$ – the gamma function for $a \geq 0$

$\Gamma(a, x) = \int_{z=x}^\infty z^{a-1} \exp(-z) \mathrm{d}z$ – the incomplete gamma function for $a \geq 0$ and $x > 0$

$\mathrm{gamma}(\beta, \lambda)$ – gamma distribution with shape parameter β and scale parameter λ

$\Gamma(\beta)^{-1} \lambda^\beta x^{\beta-1} \exp(-\lambda x) 1_{(0,\infty)}(x)$ – gamma density

$\mathrm{beta}(\beta_1, \beta_2)$ – beta distribution with parameters β_1 and β_2

$b_{\beta_1,\beta_2}(x) = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} x^{\beta_1-1}(1-x)^{\beta_2-1} 1_{(0,1)}(x)$ – beta density

$\mathrm{unif}(a, b)$ – uniform distribution on $[a, b]$

hyp-2 – hyperexponential-2 distribution (discrete mixture of two exponentials)

E$\pm n = 10^{\pm n}$, $n \in \mathbb{N}$ – power of 10 notation to represent small and large numbers

Chapter 1

Introduction

1.1 Motivation

Complex engineering systems such as bridges, roads, flood defence structures and power pylons play an important role in our society. Bridges and roads enable the transportation of goods and people from A to B, flood defense structures such as dikes and dams protect areas below sea-level from floods, and finally, power pylons facilitate the transportation of electricity from power generators to houses and companies. Unfortunately all these systems are subject to deterioration[1], meaning that in course of time their condition falls from higher to lower, and possibly even to unacceptable, levels. Obviously, the condition level and the performance of the system are related. Generally the lower the condition level the worse the performance of the system. Maintenance actions such as inspection, local repair and replacement should be done to retain the system in or restore it to an acceptable operating condition. After all, the economic consequences of malfunctioning engineering structures can be huge as the following examples show.

A Dutch bridge that is part of the highway between Amsterdam and Almere was recently (April 27, 2007) closed to truck traffic, because the construction did not meet the safety requirements (Terbruggen, 2007). Specifically, heavy vehicles could cause holes in the asphalt road and these would create unsafe driving conditions. The problem is that the state of the bridge is worse than expected, its real strength is indistinct and further deterioration is anticipated. Transport companies in the region were seriously affected by the decision to close the bridge to trucks, because they now had to take a much longer route. The total cost of this decision was estimated at 160 000 euros a day (Terbruggen, 2007). An alternative connection by ferryboat was only realized several months later (September 17, 2007), but it was not free of charge.

[1]Instead of deterioration the term degradation is also often used in the literature. As far as the author knows it has the same meaning as deterioration and so we will only speak of deterioration.

The I-35W Mississippi river bridge in Minneapolis which collapsed on August 1 2007, had already been reported 'structurally deficient' in the early 90's (O'Connell *et al.*, 1991) and 'in possible need for replacement' in 2005 (Browning, 2007). Unfortunately, the bridge did not receive appropriate maintenance and it collapsed. Apparently there was insufficient knowledge about the effects of deterioration on the bridge's structural integrity. Apart from having a big impact on the Minnesota economy, this disaster killed thirteen people.

The above examples stress the importance of regularly measuring the system condition and predicting system failure - the key elements of 'condition monitoring'. Moreover, they lead to questions such as

- How can we best model and predict deterioration?
- How to model the effect of maintenance on the system condition?
- When should which maintenance action be done?

These questions were addressed in the framework of the Dynaform project (Dynaform consortium, 2002), a research collaboration between Erasmus University Rotterdam, SemLab and Rotterdam Painting Consultants (RPC). The project ran from September 2002 until August 2004 and was sponsored by Senter, an agency of the Dutch Ministry of Economic Affairs. Maintenance and decision support knowledge was integrated in a software environment leading to an innovative maintenance management tool for complex engineering systems. A prototype of this decision support system was validated with a case study on maintenance of power pylons. The work in the project clearly indicated a number of open questions which were later addressed in this thesis. For instance, it laid out the motivation for an in-depth study of deterioration of steel structures.

Modelling deterioration

As maintenance decisions are often based on (predictions of) the condition of the system, it is important to model deterioration. In particular, the temporal uncertainty associated with the evolution of deterioration should be modelled (Pandey *et al.*, 2007). Several types of deteriorating systems can be distinguished, viz. systems where the functioning has only two levels (yes or no; intermediate condition levels cannot be measured), systems where deterioration can be classified on an ordered categorical scale and systems where deterioration is measured on a continuous (such as interval or ratio) scale. Separate stochastic models are available for these different types of deteriorating systems. Below we discuss these models.

Many electrical systems such as lamps simply function or do not function. Deterioration can then be appropriately modelled as a 0-1 state process, where 0 denotes that the system is in working condition and 1 that it has failed. Typically, system failure is

modelled by imposing structure on its lifetime or its hazard rate, the latter being the instantaneous probability of failure.

When the system condition is classified into several categories, e.g. on the basis of visual inspections, deterioration is best be modelled by (semi-)Markov chains. Instead of only 1 transition from 'working' to 'failed', Markov chain models allow for transitions between a finite number of intermediate states. Modelling this type of deterioration can thus be done by imposing structure on the state transition probabilities. In his thesis Kallen (2007) applies Markov chain models to analyze bridge deterioration, which is classified on a scale from 0 to 5 in the Netherlands.

In some systems deterioration can be measured on a continuous scale. For instance, the surface hardness of rotating systems such as ball-bearings can be measured to observe signs of surface fatigue. Furthermore, many civil infrastructure systems are subject to (slow) deterioration such as wear, creep, corrosion and crack growth. These quantities can also be measured on a continuous scale.

One of the problems tackled in the case study on maintenance of power pylons in the Dynaform project is modelling the deterioration of such structures. Generally zinc and organic coatings are applied to protect power pylons against steel corrosion. The deterioration of steel structures is measured (inspection practice) on a scale from 0 (failed) to 100 (as-good-as-new). For maintenance planning it is important to make a prediction of the deterioration in time. In principle one should do a very extensive statistical research to compare deterioration models. However, data on the state of the system was not available. Therefore, a method was developed to estimate the deterioration process via expert judgement. This method was used to estimate various models and can be found in Chapter 2. Motivated by this example we will focus in this thesis on deterioration that can be expressed by some continuous measurable quantity indicating or relating to failures. This type of deterioration will be referred to as *measurable deterioration*.

Note that before one can model measurable deterioration one should identify the different deterioration mechanisms. If there is more than one deterioration mechanism then one can either define a virtual component for each of the mechanisms or define a performance indicator to combine the individual deterioration mechanisms to a single summary scale. Secondly, one should know how to measure deterioration, e.g. by visual inspection or by means of specific measurement equipment. In the models presented in this thesis the deterioration mechanisms and the way deterioration is measured are assumed to be already established, although the latter is sometimes a difficult technical problem.

In the literature on time-dependent measurable deterioration two classes of probabilistic models can be distinguished, viz. the random variable models and the stochastic process models (Pandey *et al.*, 2007). The former are regression-type models with randomized parameters capturing the sampling variability observed in e.g. a sample of

deterioration data of identical components. Examples of these models can be found in the book by Meeker and Escobar (1998). On the other hand, stochastic process models take into account the temporal uncertainty associated with the evolution of deterioration. This is especially important when dealing with systems subject to outdoor weathering conditions, because dynamic environments may induce changes in the physics of failures (Singpurwalla, 1995).

Models based on stochastic processes such as Brownian motion (see e.g. Doksum and Hóyland, 1992; Whitmore and Schenkelberg, 1997), the gamma process (see e.g. Çinlar *et al.*, 1977; Van Noortwijk and Klatter, 1999) and the compound Poisson process (see e.g. Sobczyk, 1987) have been successfully applied to describe and predict deterioration such as wear and corrosion. These models all belong to the class of Lévy processes, which have stationary and independent increments. The stationarity implies that the deterioration increments are identically distributed and hence the expected deterioration is linear in time. However, deterioration processes such as creep, sulfate attack and scour-hole development are non-stationary (Van Noortwijk, 2007). Such deterioration can best be modelled by applying a time-transformation to one of the above-mentioned Lévy processes. This is exactly what was done in the Dynaform project to model the deterioration of coatings on steel structures. In that project we encountered the following questions.

Firstly, the question arises which of these processes is the best deterioration model. To answer this question we will compare, both theoretically and empirically, Brownian motion with a time transformation and the gamma process with non-linear shape function. Such a comparison has never been made. Some authors have pointed out that Brownian motion type deterioration models are not suitable because they are not monotonic in that the increments can also be negative. On the other hand, the analysis of these models is relatively easy (due to the normal distribution). We will study whether this advantage outweighs the disadvantage of having negative increments. A related question which we will also study is "Do these deterioration models yield similar maintenance decisions?".

Secondly, as the deterioration of engineering systems has a physical cause, would it not be better to model the underlying physical process itself? This question can only be answered if we also consider a relatively unknown class of deterioration models, namely the white-box or physical process models. As opposed to the stochastic process models, referred to as grey-box models, in white-box models structural insights in the underlying physical process are incorporated. This gives the white-box models a clear advantage over the stochastic process models. The reason that these models have always got less attention in the statistical analysis of deterioration is that conventional approaches to statistical inference are not feasible due to the mathematical intractability of the white-box models. For instance, applying likelihood methods is not feasible since computing the likelihood becomes too complex. So the question is, how to overcome this problem? In this

thesis we will present a likelihood-based framework for statistical analysis of white-box deterioration models, in which the mathematical intractability of the white-box model is circumvented by using computer intensive methods such as simulation, density estimation and stochastic approximation. Moreover, in an empirical study we compare the performance of a physical process model for steel corrosion with the performance of Brownian motion and gamma process models.

Modelling the effect of maintenance

The next issue is to model the effect of maintenance on the condition of the system. In particular, we are interested in maintenance that does not restore the condition of the system to as new, i.e. 'imperfect maintenance' or 'partial repair'. This interest is again motivated by the maintenance of coatings on steel structures. Here, maintenance actions such as 'spot repair' and 'repainting' do not reduce the deterioration of the steel structure to 0, because not all corrosion is removed. For maintenance planning purposes the size of this partial reduction is important and hence it should be modelled.

Van Noortwijk and Frangopol (2004) give a mathematical description of the lifetime extending maintenance (LEM) model, which was first proposed by Bakker *et al.* (1999). In this model it is assumed that the deterioration of a new system follows a gamma process and the effect of imperfect maintenance is twofold. Firstly, after maintenance the deterioration level is decreased by a fixed amount and, secondly, the rate of deterioration can either be 'repeating' or 'non-repeating'. Repeating implies that the rate of deterioration equals the rate at the initial condition. Non-repeating means that it equals the rate it had just before maintenance.

In order to model the effect of coating maintenance actions such as spot repair and repainting, we need to extend the LEM model. Firstly, spot repair and repainting do not cover all corrosion, as not all may be visible, and hence the reduction in deterioration is a random variable. Secondly, the above imperfect maintenance actions only improve the condition locally and hence the deterioration process increases at a different rate after maintenance. To describe this behaviour we will present a model where the reduction in condition is a random variable and the parameters of the deterioration process can be adjusted after a maintenance action.

In this model the time between two maintenance actions is the first time a gamma process exceeds a random threshold. Such a randomized hitting time was first introduced in the field of structural reliability by Abdel-Hameed (1975), who showed that this hitting time inherits the characteristics of the random threshold. The question remains how to efficiently compute the probability distribution of the randomized hitting time, being a two-dimensional integral. This is especially important in optimization models, where it is necessary to compute this distribution repeatedly. In Chapter 5 of this thesis we study

this computational issue.

Modelling maintenance decisions

Once deterioration has also been modelled in the presence of imperfect maintenance, one is ready to determine when to do which maintenance action. This is the actual maintenance optimization step. In general, the decision when to do maintenance can be time-based or condition-based. In the former case maintenance is done at prefixed points of time, in the latter case the maintenance timing is based on deterioration measurements such as inspections. The aim of maintenance optimization models is to find the set of maintenance times and actions that optimizes some criterion. Decision criteria can either be cost-based, e.g. the expected average maintenance or operating cost per unit time, or availability-based, e.g. the long-term fraction of time the system is in operating condition.

In a maintenance optimization model the deterioration of the system under consideration is related to the system lifetime. When dealing with measurable deterioration, one can set an intervention level on the amount of deterioration and define the lifetime of the system as the event in which the deterioration exceeds the intervention level[2]. A corrective replacement can then be done to restore the system to as new, but this is generally more expensive than a preventive replacement, which is done before the intervention level is exceeded.

Examples of maintenance optimization models for systems subject to *stationary* stochastic measurable deterioration are given by Abdel-Hameed (1987); Park (1988a,b); Kong and Park (1997). In these models periodic inspections are done to reveal the condition, represented by the gamma process, of the system. The system is renewed upon failure or when an inspection reveals that a preventive replacement level is exceeded. That is, both maintenance actions are perfect. The objective is to find the preventive replacement level and inspection frequency that minimize the expected average maintenance cost per time unit. Newby and Barker (2006) have extended the above model with imperfect repair and give exact solution procedures when deterioration is given by a stationary gamma process. Castanier *et al.* (2003) consider a similar model.

As we encountered non-stationary deterioration in the Dynaform project, we need to extend the above models. Hence, we will present an optimization model for imperfect maintenance (such as spot repair or repainting), where the deterioration is modelled by a non-stationary gamma process. Maintenance is done as soon as a fixed intervention level is exceeded, i.e. we assume the system is continuously monitored. Maintenance does not only reduce the condition by a random amount, but it may also influence the parameter values of this process. The objective is to find the set of (imperfect) maintenance actions

[2]We assume here that the intervention level is set from outside, e.g. due to regulations. Setting the intervention level is a very case specific affair and hence we will not dwell upon.

that minimizes the expected maintenance costs over a finite horizon.

Multi-component systems
Finally, in the Dynaform project we came across the problem of how to schedule the maintenance of an entire line of power pylons efficiently. One of the issues here is that a power pylon consists of many components and for each component a different maintenance policy may apply. However, the optimal individual policies may not be the joint optimal solution. By combining or grouping maintenance activities costs may be saved, especially when components are identical and require the same type of maintenance action (economies of scale). A complicating factor is that the components of power pylons are dependent in that some components cannot be worked on at the same time, as this would imply that the power supply has to be switched off.

To get acquainted with optimization models for the maintenance of multi-component systems we have written a survey of the literature on this topic. This study is included at the end of this thesis. It builds upon the three main dependencies in these systems: economic, stochastic and structural dependence. In the Dynaform project we used the insights obtained from the survey to develop some heuristic methods, but more research is needed to publish them.

1.2 Outline

The basis of this thesis are six papers on topics in maintenance optimization. They are all motivated by the research problems encountered in the Dynaform project. We do not pretend to have found full scientific solutions to all these problems, but we will present methods for (i) estimating and modelling deterioration, (ii) modelling and optimizing imperfect maintenance and (iii) optimizing maintenance of multi-component systems. The papers are self-contained and can thus be read independently of each other.

In the first three chapters the focus is on deterioration modelling. In Chapter 2 we study the deterioration of coatings on steel structures and to this end we use, and implicitly compare, three stochastic deterioration models, viz. Brownian motion with a time-transformation, the gamma process with non-linear shape function and the simulation of a physical process. The former two models are well-known in the literature, but as far as we know an empirical comparison has never been made. The simulation of a physical process is a white-box model in that it is based on the physical aspects of deterioration, i.e. initiation and propagation. All models are estimated from inspection data and data obtained via expert judgement for two existing steel structures in the Netherlands. This chapter has been published in *Reliability Engineering & System Safety*.

White-box deterioration models can be represented by discrete event systems, dynamic systems that result in the occurrence of events at particular points in time; each event has a certain discrete type. In Chapter 3 we introduce a framework for statistical inference on discrete event systems and illustrate the framework with an application to modelling steel corrosion. This chapter is inspired by the white-box model presented in Chapter 2. The contribution is that the proposed framework facilitates the statistical analysis of discrete event systems such as white-box deterioration models, which are often hard to analyze and as a result have not got much attention. The white-box approach allows to capture structural insights in the underlying deterioration process, e.g. by incorporating environmental effects as covariates. This chapter is based on Nicolai and Koning (2006).

In Chapter 4 a thorough theoretical comparison of Brownian motion and gamma process based deterioration models is made. Although Brownian motion type models also allow for negative increments, i.e. they lack the desired monotonicity property, many studies rely on such models. We answer the question whether Brownian motion is a proper model for (monotone) deterioration. We compare both models with respect to lifetime as well as their implications on maintenance decisions. This chapter is based on Nicolai and Dekker (2007a).

The next challenge is modelling and optimizing imperfect maintenance for systems subject to non-stationary deterioration. The deterioration model for a new system can be considered input at this stage. Based on the results obtained in the first three chapters, we choose to model the deterioration of a new system by a non-stationary gamma process. In Chapters 5 and 6 we model the effect of imperfect maintenance on deterioration and we present an optimization model which can be employed to determine when to do which maintenance action.

We derive in Chapter 5 several results for non-stationary gamma processes that are required for the optimization of imperfect maintenance. In particular, we focus on the probability distribution of the first time a non-stationary gamma process exceeds a random threshold, being a randomized hitting time. In general it is time-consuming to compute its distribution. Hence we present two approximations and study whether they are fast and reliable. This chapter is based on Frenk and Nicolai (2007).

In Chapter 6 we present an optimization model for imperfect maintenance. Maintenance is to be done as soon as a fixed threshold is exceeded. The direct effect of imperfect maintenance is considered random and hence the time between two imperfect maintenance actions is a randomized hitting time. Moreover, the indirect effect is modelled by allowing the gamma deterioration process to have different parameter values after imperfect maintenance. We focus on steel structures protected by coating systems, for which three (imperfect) maintenance actions are available: local spot repair, repainting and replacement of the coating. Finally, a dynamic programming model is developed to find the

set of actions that minimizes the expected maintenance costs over a finite time-horizon. This chapter is based on Nicolai, Frenk, and Dekker (2007b).

Finally, Chapter 7 presents an overview of maintenance models for multi-component systems. We distinguish between models for systems with economic (either positive or negative), structural and stochastic dependence between components. These dependencies all have their own effect on the type of maintenance policy applied to the system. In systems with positive economic dependence, it is beneficial to group similar maintenance activities due to economies of scale. On the other hand, negative economic dependence may be a reason to stick more to an individual planning for each component, as grouped maintenance may lead to costly breakdowns. In systems with stochastic dependence maintenance policies that combine preventive and corrective maintenance activities are often applied, because the breakdown of one component affects the failure probability of other components. Structural dependence applies if components structurally form a part, so that maintenance of a failed component implies maintenance of working components.

The contribution of this chapter is threefold. The fact that there are so many different possible dependencies between components asks for a classification in terms of these dependencies. Therefore, we present a new classification scheme for multi-component maintenance models which is founded on the three above-mentioned dependencies. In earlier overviews the focus was only on the maintenance policies, whereas we relate the policies to the dependencies. Secondly, in earlier overview articles stochastic dependence did not receive much attention and hence we give a detailed description of the progress in this category. Moreover, in all three categories many new contributions have been published since the last survey in 1996. Therefore we think it is useful to present the trends and open areas in the field of multi-component maintenance. This chapter is an extended version of Nicolai and Dekker (2007b).

In Chapter 8 we summarize the outcomes of the studies, answer the above research questions and draw conclusions.

Chapter 2

A comparison of models for measurable deterioration: an application to coatings on steel structures*

Abstract

Steel structures such as bridges, tanks and pylons are exposed to outdoor weathering conditions. In order to prevent them from corrosion they are protected by organic coating systems. In this study we model the deterioration of the organic coating layer that protects steel structures from corrosion. Only if there is sufficient knowledge of the condition of the coating on these structures, maintenance actions can be done in the most efficient way. Therefore the course of the deterioration of the coating system and its lifetime, which is of importance for taking maintenance decisions, have to be assessed accurately. In this chapter three different stochastic processes, viz. Brownian motion with a non-linear time transformation, the gamma process with non-linear shape function and a two-stage hit-and-grow physical process, are fitted to two real data sets. In this way we are the first who compare the three stochastic processes empirically on criteria such as goodness-of-fit, computational convenience and ease of implementation. The first data set is based on expert judgement; the second consists of inspection results. In the first case the model parameters are obtained by a least squares approach, in the second case by the method of maximum likelihood. A meta-analysis is performed on the two-stage hit-and-grow model by means of fitting Brownian motion and gamma process to the outcomes of this model.

*This chapter is based on Nicolai, Dekker and Van Noortwijk (2007)

2.1 Introduction

In this chapter we consider the modelling of the deterioration process of coating systems protecting steel structures, which are exposed to outdoor weathering conditions. The aim of the deterioration modelling is to predict the quality of the coating in time in order to determine when inspections or maintenance should be done and to assess life cycle costs. In general, deterioration can be modelled in several ways:

- as a black-box statistical time to failure (such as lifetime distribution);
- as a grey-box stress-strength model based on a measurable quantity indicating time-dependent deterioration and failure (such as Brownian motion and gamma process);
- as a white-box model through simulation of the physics of measurable deterioration and failure.

The latter two cases can be split up into cases where the condition of the system is monitored continuously and cases in which inspections have to be performed. In this chapter, we consider the latter case, because inspections have to be done to reveal the condition of the coating. Moreover, we consider the case where deterioration can be expressed through a measurable quantity for which failure/intervention limits can be set.

In the literature on deterioration modelling, both the gamma process and Brownian motion have been used to predict the value of a measurable quantity. Dekker *et al.* (1998a) use a Brownian motion process to describe different forms of deterioration of asphalt roads. Van Noortwijk and Klatter (1999) model the scour-hole development in block mats as a gamma process. Van Noortwijk (1996) also uses the (generalised) gamma process to describe sand erosion, crest-level decline and long shore rock transport. Although Dekker *et al.* (1998a) discuss the (sometimes undesirable) property of negative increments of Brownian motion, they do not compare this process with a gamma process. On the other hand, in Van Noortwijk and Klatter (1998) and Van Noortwijk (1996) the possible disadvantages of the gamma process are not mentioned, and this process is not compared with Brownian motion.

Another way of modelling deterioration is given in Meeker and Escobar (1998). They consider the deterioration of different units, measured at different points in time (resulting in a deterioration path for each unit). They propose a regression model, where the regression function depends on (transformed) time. The regression parameters may be random and each unit may have different parameters. The error term is assumed to be normally distributed. However, it should be noted that a regression model cannot capture temporal variability associated with the evolution of deterioration. As a consequence, the deterioration along a specific sample path is deterministic in the regression model, whereas

it varies probabilistically in the Brownian motion and gamma process models (Pandey *et al.*, 2007).

A number of papers has been written on (modelling) deterioration of systems exposed to outdoor weathering conditions. For example, Chan and Meeker (2001) relate deterioration to environmental factors, such as the weather. These factors are transformed into a deterioration rate. A time series modelling approach is proposed to predict daily deterioration. Heutink *et al.* (2004) describe how the maintenance methodology used in the Netherlands is applied to protective paint systems. The lifetime extending maintenance model (LEM; Van Noortwijk and Frangopol, 2004), in which deterioration is modelled by a gamma process with expected deterioration non-linear in time, is applied successfully to optimize maintenance of the coating of the Haringvliet storm-surge barrier.

In this chapter we do not only model the deterioration of the coating on two specific steel structures by a gamma process and a Brownian motion model; we also simulate a physical process referred to as the two-stage hit-and-grow process. The former two stochastic processes are both in the class of Lévy processes. Restricting this class to processes with continuous sample paths leads to the Wiener process or Brownian motion. On the other hand, forcing monotonicity leads to jump processes (Van Noortwijk, 2007); the gamma process is a jump process. The simulation of a physical process is likely to give more insight and may perform well since it is more related to the real physical process. Because environmental deterioration processes are slow, data is scarce. This research is an extension of the methodology in an earlier study (Nicolai *et al.*, 2004). In contrast to the earlier study, all processes are now fitted to data in two forms: data based on expert judgement and inspection results. By doing so, we compare these processes empirically and determine the advantages and disadvantages for each model. Moreover, we consider the appropriateness of the models as well as the applicability for both types of data, viz. expert judgement and inspection data. The estimation of the parameters of the three processes is discussed in detail. Finally, we do not only show how to model deterioration of coating systems, but we also give detailed results of this modelling. Note that although this chapter has a clear focus on the deterioration of coating systems, the described methods should be regarded as generally applicable.

The outline of this chapter is as follows. In Section 2.2 we clarify the problem. In Section 2.3 the two different data sets are introduced. The first is based on expert judgement; the second consists of inspection results. In Section 2.4 we present three stochastic deterioration models. The topic of Section 2.5 is the parameter estimation for these models. In Section 2.6 the results of the estimation procedures are given. Finally, conclusions are drawn in Section 2.7.

2.2 Problem description

We study the problem of modelling the deterioration process of the coating on two specific groups of steel objects in the Netherlands. More information on these objects is given in Sections 2.3.1 and 2.3.2. In order to get a better understanding of the deterioration process, we shall first discuss how steel objects are protected against deterioration in general. The steel base of an object is protected against steel corrosion using a double (duplex) protection coating. The steel is galvanized, i.e. it is covered with a zinc coating that acts as a sacrificial anode. The zinc corrodes first, protecting the steel using the method of cathodic protection[1]. However, in order to prevent a too rapid corrosion of the zinc layer, which would leave the steel unprotected after several years, an organic coating protects the zinc. An intact organic coating shields the zinc from the corrosive environment and prevents it from corroding. Unfortunately, the organic coating is subject to deterioration itself. Under the influence of humidity and UV radiation, the protective properties diminish slowly in time, eventually resulting in corrosion of the zinc layer. Too much zinc corrosion is not allowed, because it releases zinc to the environment, causing pollution. Ultimately, the steel corrodes and the structural integrity of the object can be compromised.

Modelling the deterioration of coating systems is of importance when taking inspection and maintenance decisions as well as when determining life cycle costs. Not only the amount of deterioration, but also the location of the deterioration (read: corrosion) should be taken into account when planning maintenance. A widespread deterioration is much more difficult to address than a concentrated one. Black- and grey-box processes do not take the location of 'spots' into account, but processes relating more to the physics of the deterioration do. For a decision support system for maintenance optimization a visualization of the deterioration process is valuable. On the other hand, black- and grey-box processes may be more convenient to express lifetime statistics of the coating and to allow all kind of optimizations. In this chapter, we therefore also answer the question whether black- and grey-box processes can replace physical processes in a decision support system.

A number of factors have an effect on the deterioration process of a coating system. We distinguish between the environment, the type of coating system, the quality of application (of the coating) and the complexity of the object. The environment (covering local weather conditions and the atmospheric composition) and the type of a coating system are judged to have the biggest effect on deterioration. According to ISO standard $12944 - 2$ (ISO/EN 12944, 1998) the environments of the two objects we consider are classified as 'very high marine', since they are situated nearby sea. This environment is

[1]The sacrificial anode is electrically more active than the steel and therefore, the corrosive current will exit from the zinc rather than the steel. The steel is protected while the zinc is sacrificed.

the most severe with respect to the effect on deterioration of coating systems. This is the reason why these objects are both protected by thick coating systems. With respect to quality of application, experts rate the steel objects discussed in Section 2.3.1 as 'good' on a scale 'bad - mediocre - good - very good' and the steel objects discussed in Section 2.3.2 as 'mediocre / good'. The complexity of the objects is classified as 'complex' on a scale 'very simple - simple - complex - very complex'. So, the deterioration factors are quite similar for both objects, except for the quality of application.

Visual inspections are used to determine the amount of corrosion deterioration. The amount of deterioration can be expressed as the percentage of corroded surface or, alternatively, as the percentage of surface to be repaired. Corrosion usually starts in small spots which grow in time giving rise to a more-or-less random pattern of spots on a whole surface. Determining the actual percentage of corroded surface of large objects by visual inspection is not easy as one has to do a detailed counting. It is much easier to give a rough estimate of the amount of surface to be repainted. In this study consider both deterioration measures and using a simulation model we can link them. There also exist yardsticks to convert one measure to the other one. In both cases the amount of deterioration is expressed on a scale from 0 to 100%. A new coating takes value 0%; an old coating takes value 100%.

2.3 Data

In Sections 2.3.1 and 2.3.2 we discuss the two different data sets. In Section 2.3.3 we mention the implications that the different forms of data have on modelling deterioration.

2.3.1 Expert judgement

We asked experts in the field of painting deterioration to specify the time in years in which they think that the percentage of the coating surface to be repaired (on a specific steel object) exceeds 5%, 15% and 30%. The answers they provided are given in Table 2.1. Note that they did not specify point estimates but intervals. The experts did not specify the probabilities of exceeding the given deterioration levels in these intervals.

Deterioration level	Interval in years	Numbering
5	$10-11$	1
15	$13-14$	2
30	$15-17$	3

Table 2.1: Data obtained by expert elicitation.

In order to find a relation between the deterioration level and time we first fitted a number of deterministic curves to the data in Table 2.1. The middle point of each time interval was chosen as a point estimate for the time that the corresponding deterioration level is exceeded. This way it is possible to apply least squares fitting. We fitted two functions to the data: a power law function of the form μt^q and a logistic function of the form $100(1-\exp(-\lambda t^\gamma))$. With respect to the choice of the first function we can say that many physical relationships can be modelled by a power function. The latter function is appropriate when the amount of deterioration is bounded or approaches a saturation point, since it can produce an s-shaped curve. Note that the amount of deterioration is bounded by 100% in this case. However, we are mainly interested in deterioration up to 30% (expert judgement), since maintenance should be done when this level is exceeded. Deterioration will only slow down later. So we actually may expect that the power function gives a good approximation of the expected course of deterioration.

It appears that the power function gives a better fit than the logistic function. The errors made by the logistic function are 5 times higher than the errors made by the power function. The parameter estimate of the power q, 4.16, suggests that the deterioration is non-linear in time. Therefore in the next section we assume that the expected value of the stochastic processes follows a non-linear power function.

2.3.2 Inspection data

The inspection data are taken from Heutink *et al.* (2004) and concern the coating on the steel gates of the Haringvliet storm-surge barrier in the Netherlands. In 1970, the sea arm Haringvliet was closed off from the sea by the Haringvliet barrier. In the period from 1988 to 1996, the old coatings were replaced by new ones. In 2002, half of these coatings were inspected resulting in different inspection intervals. The available inspection data represent the percentage of the seaside surface of a steel gate that has been corroded due to ageing of the coating. The area to be repaired is about 6.75 times larger than the corroded area. The inspection results are given in Table 2.2. It is clear that we deal with inspections from different gates, since the inspection value in year 12 is lower than the value observed in year 10.

2.3.3 Comparison of the data

In the previous sections we have seen that the experts provided us time intervals, while the inspections result in point estimates. Both forms of data have their disadvantages. Firstly, in expert judgement effects such as anchoring play a role; the estimates are influenced by the initial beliefs of the experts. On the other hand, inspection data may contain errors due to measurement errors and they are costly to obtain. Secondly, estimating the

Time (years)	Corrosion (%)
6	0.27
8	0.41
10	0.84
12	0.75
14	2.10

Table 2.2: Inspection results for the five Haringvliet steel gates.

parameters of the stochastic models from these data is done in a different way. For the expert data we will estimate the parameters of the processes by a least squares approach; for the inspection data we apply the maximum likelihood method. In Section 2.5 we return to this subject.

2.4 Stochastic deterioration models

In this section we introduce three stochastic processes that are suitable for describing the deterioration of organic coatings protecting steel structures. In Section 2.3 we have seen that the expected development of the deterioration process is non-linear in time. Therefore, we first describe deterioration by two grey-box processes with non-linear drift; a Brownian motion (BM) type process and the gamma process. Subsequently, we present a white-box two-stage physical process.

2.4.1 Brownian motion with a non-linear time transformation

Let us define $D(v)$ as the stochastic deterioration v time units after the organic coating is renewed, and $D(0) = 0$ with probability 1. If we assume that the stochastic process $D = \{D(v), v \geq 0\}$ is a Brownian motion with *linear* drift on the time scale v (see e.g. Karlin and Taylor, 1975), then we may write

$$D(v) = \mu v + \sigma W(v), \ v \geq 0. \tag{2.1}$$

In Equation (2.1) $W(v)$ is a standard Brownian motion, μ is the drift parameter and σ is the volatility parameter of the process. The mean and the variance of $D(v)$ are given by μv and $\sigma^2 v$, respectively. Note that in the class of Lévy processes, Brownian motion is the only one having continuous sample paths. Brownian motion cannot be monotone, since forcing a Lévy process to be monotone leads to a jump process (Van Noortwijk, 2007).

In many papers on modelling deterioration with a Brownian motion a time transformation is proposed, see e.g. Doksum and Hóyland (1992), Whitmore and Schenkelberg (1997), and more recently, Wang and Nair (2005). Doksum and Hóyland (1992) were the first to propose the time transformation in its general form. Real time t is transformed to time $v = \Lambda(t)$, where Λ is a non-decreasing function of time t. In order to model variable levels of stress in an accelerated life time experiment, they apply a time transformation which is linear with respect to time t. Whitmore and Schenkelberg (1997) consider an exponential time transformation $v = 1 - \exp(-\lambda t^{\gamma}$ and a power time transformation $v = t^q$ for modelling the deterioration of self-regulating heating cables. In Doksum and Hóyland (1992) and Whitmore and Schenkelberg (1997) the parameters are estimated by applying the method of maximum likelihood. On the other hand, Wang and Nair (2005) introduce a non-parametric method to estimate (the parameters of) the time transformation function Λ, without any assumption on its form.

We now apply the power time transformation $v = t^q$. This transformation is suggested by the prevalence of power relationships in physical models generally (see e.g. Çinlar *et al.*, 1977, Sobczyk and Spencer Jr, 1992, Whitmore and Schenkelberg, 1997, and Newby, 1998). Moreover, it is motivated by the preliminary results mentioned in Section 2.3.1. We expect that $q > 1$, which implies that deterioration is accelerating in time. Applying this transformation to our stochastic process, results in a Brownian motion type process with non-linear drift on the time scale t (measured in years). A disadvantage of this process is that it has not only positive but also negative increments. This does not seem to be realistic since the deterioration level cannot decrease in practice, unless there are errors within the observations. However, there are no big negative increments as long as the drift is large (i.e. parameters μ and σ are large) and the variance is small (i.e. σ is small). Moreover, an occasional improvement is not problematic since maintenance decisions usually depend on observed deterioration, which rarely improves if the time between two inspections is substantial (Dekker *et al.*, 1998a). An advantage of this process is that the increments are normally distributed and this implies that the process itself is normally distributed at every point in time. Therefore, computing the distribution of $D(v)$ and simulating the process are easy.

Let us first look at some properties of Brownian motion with drift defined by Equation (2.1) and assume that $v = t$. Let $T_D(z)$ be the first time BM with linear drift exceeds $z > 0$, that is $T_D(z) := \inf\{t \geq 0 : D(t) > z\}$. Now the cumulative distribution function (CDF) of this hitting time is found as (Karlin and Taylor, 1975, Whitmore and Schenkelberg, 1997):

$$\begin{aligned} F_{BM}(t,z) &:= \mathbb{P}\{T_D(z) \leq t\} \\ &= \Phi\left(\frac{\mu t - z}{\sigma\sqrt{t}}\right) + \exp\left(\frac{2\mu z}{\sigma^2}\right)\Phi\left(\frac{-\mu t - z}{\sigma\sqrt{t}}\right). \end{aligned} \tag{2.2}$$

Here Φ is the CDF of a standard normally distributed random variable. Note that $T_D(z)$ follows the inverse Gaussian distribution with parameters $z\mu^{-1}$ and $z^2\sigma^{-2}$.

Let us now consider the time transformation in its general form, i.e. $v = \Lambda(t)$. If $T_\Lambda(z)$ is the hitting time of the process $D(\Lambda(t))$, then Doksum and Hóyland (1992) have shown that $T_\Lambda(z)$ and $\Lambda^{\leftarrow}(T_D(z))$ follow the same distribution. Here, $\Lambda^{\leftarrow}$ is the inverse function of Λ. Put differently, the CDF of $T_\Lambda(z)$ can be written as

$$\mathbb{P}\{T_\Lambda(z) \leq t\} = \mathbb{P}\{T_D(z) \leq \Lambda(t)\}.$$

So, Equation (2.2) can be used to compute the distribution of $T_\Lambda(z)$ with $\Lambda(t) = t^q$. Moreover, it can be used to calculate the probability that a certain level is exceeded for the first time in a certain interval in terms of the model parameters. For example, if we denote by $p^z_{t,BM}$ the probability that the deterioration exceeds a value z for the first time between years and $t-1$ and t, then

$$p^z_{t,BM} := F_{BM}(t^q, z) - F_{BM}([t-1]^q, z),\ t \geq 1. \tag{2.3}$$

If we know this probability for different combinations of v and z, then we can estimate the parameters μ, σ and q.

2.4.2 Gamma process with non-linear shape function

As far as the authors know, Abdel-Hameed (1975) was the first to propose the gamma process as a proper model for deterioration occurring random in time. A recent overview of the application of the gamma process in maintenance is given by Van Noortwijk (2007).

Let $X(t)$ be the deterioration level at time $t > 0$ and $X(0) = 0$ with probability 1. Also, let the probability density function of $X(t)$ be given by

$$f_{X(t)}(x) = GA\left(x\middle|\frac{\mu^2 t^q}{\sigma^2}, \frac{\mu}{\sigma^2}\right), \tag{2.4}$$

where $GA(x|a,b)$ is the probability density function of the gamma distribution with parameters a and b. If we assume that the stochastic process $X = \{X(t), t \geq 0\}$ has independent increments, then X is a non-stationary gamma process with mean and variance non-linear in time, i.e. μt^q and $\sigma^2 t^q$, respectively (see e.g. Van Noortwijk and Klatter, 1999). Alternatively, X is a gamma process with non-linear shape function non-linear shape function $\frac{\mu^2 t^q}{\sigma^2}$ and scale parameter $\frac{\mu}{\sigma^2}$.

The gamma process X has the same mean and variance as the above-defined BM with a time transformation $\{D(t^q), t \geq 0\}$. However, in contrast to BM, the increments of the gamma process are always positive. Actually, the gamma process is a jump process. Jump processes follow from Lévy processes by assuming monotonicity (see e.g. Van Noortwijk,

2007). As a consequence, the CDF of the hitting time can be easily derived from the probability distribution of itself. We define $T_X(z)$ as the first time the process exceeds a level $z > 0$, that is

$$T_X(z) := \inf\{t \geq 0 : X(t) > z\}.$$

Denoting by $F_{GA}(t, z)$ the probability that the first time the gamma process exceeds z is before time t, we have that (see e.g. Van Noortwijk and Klatter, 1999)

$$F_{GA}(t, z) := \mathbb{P}\{X(t) > z\} = \int_{x=z}^{\infty} f_{X(t)}(x)\mathrm{d}x = \frac{\Gamma\left(\frac{\mu^2 t^q}{\sigma^2}, \frac{\mu z}{\sigma^2}\right)}{\Gamma\left(\frac{\mu^2 t^q}{\sigma^2}\right)}. \tag{2.5}$$

Here, $\Gamma(a, x)$ is the incomplete gamma function for $x \geq 0$ and $a > 0$, and $\Gamma(a)$ is the gamma function for $a > 0$. Now, we can compute $p^z_{t,GA}$, the probability that the gamma process X exceeds z for the first time between years $t - 1$ and t:

$$p^z_{t,GA} := F_{GA}(t, z) - F_{BM}(t - 1, z), \ t \geq 1. \tag{2.6}$$

Note again that, as for BM, we can use this expression to estimate the model parameters μ, σ and q. Also, Equation (2.5) enables us to compute the probability distribution functions of $T_X(z)$ and $X(t)$ for $z > 0$ and $t > 0$, respectively.

Time transformation of the shape function of the gamma process

We have seen earlier that Brownian motion with non-linear drift function is derived from a standard BM by transforming the time. In a similar way we can show that the non-stationary gamma process can be derived from a stationary gamma process. Below we will first discuss this relation and derive at the same time the relation between the first hitting times of these processes. To this end we first define the gamma process $Y_{v,u} = \{Y_{v,u}(t), t \geq 0\}$ having increasing shape function v and scale parameter $u > 0$. Mean and variance of $Y_{v,u}(t)$ are then given by $u^{-1}v(t)$ and $u^{-2}v(t)$, respectively. Let us define the first time $Y_{v,u}$ exceeds level z by $T_{v,u}(z)$. It is now easy to see that $Y_{v,1}(t)$ follows the same distribution as $uY_{v,u}(t)$. Moreover, we can prove that the first time $Y_{v,1}$ exceeds uz, given by $T_{v,1}(uz)$, has the same distribution function as $T_{v,u}(z)$. This means that we may restrict ourselves to gamma processes with shape function $v(t)$ and a scale parameter equal to 1.

We will now relate the hitting time $T_{v,u}(z)$ of the non-stationary gamma process $Y_{v,u}$ to the hitting time of a stationary gamma process. Let $Y_{id,1} = \{Y_{id,1}, t \geq 0\}$ be this stationary gamma process with shape function $v(t) = t$ and scale parameter 1. Also let $T_{id}(z)$ be the first time $Y_{id,1}$ exceeds level $z > 0$.[2] Notice that $Y_{id,1}(v(t))$ follows the same

[2]Here, the subscript id indicates that the shape function is equal to the identity function, which is defined as $id(x) = x$.

distribution as $Y_{v,1}(t)$. So, applying a time transformation $v(t)$ to a stationary gamma process results in a non-stationary gamma process with shape function $v(t)$. Now it is easy to see that $T_{v,u}(z)$ and $v^{\leftarrow}(T_{id,1}(uz))$ follow the same distribution. A direct consequence of this result is that

$$\mathbb{P}\{T_{v,u}(z) \leq t\} = \mathbb{P}\{v^{\leftarrow}(T_{id,1}(uz)) \leq t\} = \mathbb{P}\{T_{id,1}(uz) \leq v(t)\},\ t \geq 0,$$

which finally relates the hitting times of the stationary and the non-stationary gamma processes.

2.4.3 Two-stage hit-and-grow process

In this section we present a model that is based on the physics of the deterioration process of coating systems. This model is related to the contact process introduced by Harris (1974). This process models interactions of sites on a lattice, where every site can either be infected or not infected. The sites on the lattice form a so-called $\{0, 1\}$-valued random field. In the contact process infected sites infect direct neighbours at a certain rate and they recover at a certain rate. Most of the work on the contact process treats the asymptotic behaviour of the process. We refer the interested reader to Chapter 7 of Bremaud (1999) for an introduction to (Markov) random fields. Probably the most well-known model in the context of Markov random fields is the Ising model, introduced by Ising in 1925. This model considers the physics of phase transitions, which occur when a small change in a parameter causes a large-scale, qualitative change in the state of a system (Cypra, 1987). We model the deterioration of a coating layer as a random field, where the coating layer is represented by a grid consisting of squares. The random field is the collection of $\{0, 1\}$-valued random variables which present the condition of a square. An infected square is called a spot and such a spot will infect its neighbours after some time. Some readers may see similarities between our model and Conway's game of life (Gardner, 1970), but in the game of life the state of the system is fully determined only by its initial state. In our model this is not the case, because spots appear on the grid randomly, in time as well as in space.

The model presented here is based on the assumption that deteriorated spots arise at random positions on the coating layer. After the appearance of such a spot it expands by 'infecting' adjacent parts of the coating layer. In reality, a spot has approximately the form of a circle and experts stated that its expansion is a quadratic function of the growth of deterioration in one direction. For convenience we model expansion by declaring squares adjacent to spots infected after some (random) time. This results in the desired quadratic expansion. According to the experts in the field of coating deterioration, modelling the expansion in this way is consistent with the way deterioration is measured in practice. An advantage of this hit-and-grow process is that it is monotonically increasing, i.e. it has

positive increments only. Brownian motion does not have this property, but the gamma process has. The main advantages of the model are

- it gives insight into the physics of the deterioration process, since it is related to the actual physical process,
- the effect of a maintenance action on deterioration can be evaluated and quantified.

Here we do not incorporate maintenance actions. Below we describe the two stages of the model. In the first stage a spot randomly hits the surface, this is called initiation; in the second stage the spot starts growing and this called propagation. We refer to this model as the two-stage hit-and-grow process. For convenience we shall use the abbreviation TSHG.

Initiation of spots

We assume that spots hit the surface according to a non-homogeneous Poisson process (NHPP) $N = \{N(t), t \geq 0\}$ with intensity rate function $\eta(t) = \lambda t^p$, where $\lambda > 0$ and $p > -1$. This means that the initiation process is non-stationary unless $p = 0$. We allow for non-stationary increments since experts stated that the initiation of spots depends on the permeability of the coating layer, which becomes more permeable as it deteriorates. So, it is likely that the intensity rate function is an increasing function of time ($p > 0$). Note that for this NHPP $m(t) := \mathbb{E}(N(t))$, the expected number of arrivals until time t is equal to $m(t) = \hat{\lambda} t^{p+1}$, with $\hat{\lambda} = \frac{\lambda}{p+1}$. Introducing now a new time scale $y = \phi(t) = t^{p+1}$, we see that the initiation of spots can be modelled as a HPP with rate $\hat{\lambda}$ on time scale y. This time transformation simplifies the simulation of the NHPP, since the interarrival times of a HPP are exponentially distributed with parameter $\hat{\lambda}$. In this case the arrival times of the NHPP can be retrieved by transforming the time back to time scale $t = \phi^{\leftarrow}(y) = y^{1/(p+1)}$.

Now imagine the coating layer as a grid consisting of N^2 unit squares, having coordinates (i, j), $i, j = 1, \ldots, N$. At the arrival time of a spot, it is 'assigned' to a random position on the grid in such a way that each position is equally likely for each spot, that is each position is chosen with probability N^{-2}. If, at a certain point in time a spot is assigned to a square which was already hit before, then its condition will remain 1, i.e. infected. In such a case we will not select another square.

Note that we can generalize the initiation phase by defining independent Poisson processes $N_{i,j} = \{N_{i,j}(t), t \geq 0\}$, which count the number of spots arrived in square (i, j) at time t. These non-homogeneous Poisson processes can be regarded as at-most-1-jump processes. The corresponding intensity functions are given by η_{ij}. The superposition of the processes N_{ij}, defined by $N = \sum_{N_{ij}}$, is again a Poisson process with intensity function equal to $\sum_{i,j} \eta_{ij}$. So, in general the probability that a spot, once it has arrived, is assigned

to square (i, j) is given by

$$\frac{\eta_{ij}}{\sum_{k,l} \eta_{kl}}.$$

Choosing the same intensity function for each square on the grid results in the probability N^{-2} mentioned above. In Chapter 3 we even consider general point processes.

Propagation of spots

The expansion (propagation) of spots is modelled as follows. After a spot has appeared on the grid it infects all adjacent squares after δ time units. Note that a square has a maximum of eight adjacent squares. Infected squares also infect their neighbours δ time units after they were infected themselves. We assume that this interexpansion time (or 'interinfection' time) is constant over time. This seems to be a realistic assumption since the expansion depends on the purity of the surface, which appears to be constant in course of time. We will also experiment with a random interexpansion time Δ following an exponential distribution with mean δ. In this case the interexpansion times are realizations of the random variable Δ. In this way, we also model the variability in the propagation of spots. We suspect that the exponentially distributed interexpansion times result in more interaction between the spots than the model with constant interexpansion time, because due to chance rapid expansions will also occur.

It may have become clear from the preceding that the analysis of the TSHG process can only be done analytically to a limited extent. Asymptotic statements can be made, but we are interested in the time-dependent behaviour. Problems arise when spots start to overlap, as in that case probability statements are difficult to make. Therefore we will rely on a simulation modelling. Every spot corresponds to a square (block) in a grid which is the smallest perceptible amount of corrosion. Notice that the fineness of the grid determines the spot size which is also related to the minimal size that can be detected by inspection. Having many squares results in a more detailed modelling, but is also much more computational demanding. For example, if we consider a sluice door of 100 m^2 and take 10 000 (1 000 000) squares in the simulation model, then each square represents an area of 100 (1) cm^2.

The fact that we consider the initiation of spots as a NHPP makes simulation easy. Simulating the propagation stage is somewhat more difficult since as more spots arise on the grid, the number of potential expansions also increases. This requires a lot of bookkeeping and, as a consequence, working memory. The simulation was implemented in the statistical software package R (R Development Core Team, 2005).

2.5 Parameter estimation

In Sections 2.5.1 and 2.5.2 the parameter estimation of the three processes from the expert data and the inspection data is discussed, respectively. In Section 2.5.3 we introduce a meta-analysis of the TSHG process. With respect to this we discuss the fitting of BM and the gamma process to simulated interval probabilities.

2.5.1 Expert data

In this paragraph we will give a method for estimating the parameters of the three stochastic processes by expert judgement. In some articles on deterioration modelling expert judgement has been used to estimate (some of) the parameters of a stochastic process. For example, Wang *et al.* (2000) model the hazard rate of water pumps in the presence of preventive maintenance (PM) as a gamma process. In the absence of PM the mean of the gamma process is taken as the hazard rate of the Weibull distribution. The scale parameter of this distribution is estimated by expert judgement. The method of maximum likelihood is applied to estimate the other parameters, such as the scale parameter of the gamma process itself and the parameters that model the effect of PM on the hazard rate. With the method introduced below all parameters can be estimated simultaneously.

We have already mentioned that the answers of the experts in Table 2.1 include uncertainty. Unfortunately, the experts did not give a good estimate of the probability that deterioration level is exceeded in interval i, for $i = 1, 2, 3$. Therefore, we assign this probability to each interval ourselves. Let us define α_i as the probability that deterioration level i is exceeded in the ith interval. Now we can estimate the model parameters by equating α_i for $i = 1, 2, 3$ with the probabilities that result from the stochastic processes. For example, estimating the parameters of the gamma process comes down to solving the following system of non-linear equations with respect to the model parameters:

$$\begin{array}{rcccl} p^{5}_{11,GA} & = & F_{GA}(11,5) - F_{GA}(10,5) & = & \alpha_1 \\ p^{15}_{14,GA} & = & F_{GA}(14,15) - F_{GA}(13,15) & = & \alpha_2 \\ p^{30}_{16,GA} + p^{30}_{17,GA} & = & F_{GA}(17,5) - F_{GA}(15,30) & = & \alpha_3. \end{array} \tag{2.7}$$

For Brownian motion the equations are similar; only the formulas for the hitting time probabilities are different. For given values of the interval probabilities α_i, $i = 1, 2, 3$, the above system can be solved for μ, σ and q. Alternatively, one can also minimise the sum of squared differences (SSD) between the left and right hand sides (LHS, RHS) of the equations with respect to μ, σ and q. We implement the latter approach and apply a standard optimization procedure to find the parameter values that minimize the SSD.

Estimating the parameters of the TSHG process is essentially the same as for the other two processes, but the difference is that this process is simulated. It appears to

be impossible to derive characteristics such as the probability distribution of the TSHG process analytically. Estimating the parameters now comes down to finding values for λ, p and δ such that in ($100\alpha_i\%$ of the simulation runs deterioration level i is exceeded in the ith interval. To this end we minimise the following stochastic objective function:

$$f(\theta) = \sum_{i=1}^{3} (G_i(\theta) - \alpha_i)^2. \tag{2.8}$$

In this equation θ is the parameter vector (λ, p, δ) and $G_i(\theta)$ is the fraction of simulation runs in which the deterioration level corresponding to the th interval is exceeded. In order to minimize (2.8) we apply a grid search procedure that evaluates (2.8) for many combinations of the parameters λ, p and δ. The stochastic objective function is evaluated by simulating the TSHG process 1 000 times on a grid of 10 000 squares, for a period of 20 years. All by all, this makes the parameter estimation a time-consuming process.

2.5.2 Inspection data

In order to fit a Brownian motion or a gamma process to inspection data, statistical methods for the parameter estimation are required. For a single component (i.e. steel gate), a typical data set consists of n inspection times t_i, $i = 1, \ldots, n$, where $0 = t_1 < t_2 < \cdots < t_n$, and corresponding observations of the cumulative amounts of deterioration x_i, $i = 1, \ldots, n$, where $0 = x_1 < x_2 < \cdots < x_n$.

The parameters μ, σ, and q of the gamma process can be estimated by maximizing the likelihood function of the independent increments of deterioration (Çinlar *et al.*, 1977) with respect to μ, σ, and q. The likelihood function of the observed deterioration increments $y_i = x_i - x_{i-1}$, $i = 1, \ldots, n$, is a product of independent gamma densities:

$$\prod_{i=1}^{n} f_{X(t_i)-X(t_{i-1})}(y_i) = \prod_{i=1}^{n} GA\left(y_i \middle| \frac{\mu^2(t_i^q - t_{i-1}^q)}{\sigma^2}, \frac{\mu}{\sigma^2}\right). \tag{2.9}$$

This likelihood function can be extended from a single component to multiple components (i.e. steel gates) by considering m independent components, $j = 1, \ldots, m$, for which n_j inspections are performed resulting in n_j independent deterioration increments, i.e.

$$\prod_{j=1}^{m}\prod_{i=1}^{n_j} f_{X(t_{ij})-X(t_{i-1j})}(y_{ij}) = \prod_{j=1}^{m}\prod_{i=1}^{n_j} GA\left(y_{ij} \middle| \frac{\mu^2(t_{ij}^q - t_{i-1j}^q)}{\sigma^2}, \frac{\mu}{\sigma^2}\right), \tag{2.10}$$

where x_{ij} is the cumulative amount of deterioration at the ith inspection time for the jth component and $y_{ij} = x_{ij} - x_{i-1j}$ is the ith deterioration increment for the jth component. In a similar manner, the parameters of Brownian motion with a time transformation can be determined by replacing the likelihood function of independent gamma distributed

increments with the likelihood function of independent normally distributed increments. For the Haringvliet steel gate data, $m = 5$ and $n_j = 1$ for $j = 1, \ldots, m$. Most statistical packages contain optimization procedures for likelihood functions. We have implemented and maximized the likelihood functions for BM and gamma process in EViews 5.0 (Quantitative Micro Software, 2004).

The TSHG process lacks the property of independent increments. This is because the number of spots on the grid determines the growth rate. The more spots on the grid, the higher the probability that another square becomes infected. This means that the likelihood function for this process can not be written as the product of the densities of the increments. We denote the density function of the deterioration $HG(t)$ at time t according to the TSHG process by $f_{HG(t)}(\cdot|\theta)$, where θ is again the parameter vector (λ, p, δ). In general, the likelihood function of the data given the model (parameters) can be written as the product of the multivariate density functions (of the components) at the inspection times evaluated at the observed deterioration values:

$$\prod_{j=1}^{m} f_{HG(t_{1j}),\ldots,HG(t_{n_j j})}(x_{1j}, \ldots, x_{n_j j}|\theta). \tag{2.11}$$

For the Haringvliet steel gate data we have $m = 5$ and $n_j = 1$, for $j = 1, \ldots, m$. Equation (2.11) then changes to

$$\prod_{j=1}^{5} f_{HG(t_{1j})}(x_{1j}|\theta). \tag{2.12}$$

The problem however is that we do not have explicit expressions for the density functions. We propose the following solution. For a given parameter vector ϑ we replace the probability density functions in Equation (2.12) by their kernel estimates. (For an introduction to density estimation we refer the interested reader to Scott, 1992). To this end we first simulate k sample paths from the TSHG process and record the realizations of the process at each of the $m = 5$ inspection times $t_{11}, \ldots, t_{1m}$. The k realizations for each inspection time are denoted by $z_{j1}, \ldots, z_{jk}$ for $j = 1, \ldots, m$. Given this sample of realizations the kernel estimate (or the Parzen-Rosenblatt estimate, Parzen, 1962) of $f_{HG(t_{1j})}(x_{1j}|\vartheta)$ is defined by

$$\hat{f}_{HG(t_{1j})}(x_{1j}|\vartheta) = \frac{1}{kh} \sum_{l=1}^{k} K\left(\frac{x_{1j} - z_{jl}}{h}\right), \; j = 1, \ldots, m. \tag{2.13}$$

The function K is a square integrable probability density (e.g. normal) and h is a smoothing parameter, usually referred to as the bandwidth parameter. A problem in kernel

density estimation is selecting the bandwidth. We follow the literature on the optimal bandwidth problem and apply the procedure proposed by Sheather and Jones (1991).

Maximum likelihood estimation now comes down to maximizing the expected value of the *simulated* likelihood function $\hat{L}(\vartheta)$ with respect to ϑ. That is, we have the following maximization problem

$$\max_{\vartheta} \mathbb{E}(\hat{L}(\vartheta)) = \max_{\vartheta} \mathbb{E}\left(\prod_{j=1}^{5} \hat{f}_{HG(t_{1j})}(x_{1j}|\vartheta)\right). \tag{2.14}$$

Considering $\hat{L}(\vartheta)$ to be the real likelihood, we can perform a gradient-free optimization procedure to find the parameter estimates numerically. Since the simulation was implemented in the statistical software package R, this package was also used to maximize the likelihood function.

2.5.3 Meta-analysis of the TSHG process

A meta-analysis is next performed on the results of the TSHG process. This is motivated by the fact that simulating the TSHG process is time-consuming. Therefore it may be better to implement a gamma process or BM in a decision support system for maintenance optimization, instead of or in conjunction with the TSHG process itself. Clearly, the gamma process or the BM should fit well to the outcomes of the TSHG process. Fitting Brownian motion or gamma process to the outcomes of the TSHG can be done in a similar way as described in Section 2.5.1. The most important difference with the estimation approach given there is that we first determine the interval probabilities by simulation (of the TSHG process). Recall that an interval probability is the probability that the deterioration process exceeds a certain level in a given interval. Another difference is that we can simulate more than 3 interval probabilities such that we have more equations than parameters. Again, the parameters are estimated by minimising the sum of the squared differences between the interval probabilities for the gamma process (or BM) and the estimated probabilities for the TSHG process.

2.6 Computational results

In this section we discuss the results of fitting the three processes to the expert data and the inspection data. We will also present the parameter estimates of Brownian motion and the gamma process when fitted to the outcomes of the (simulated) two-stage hit-and-grow process.

2.6.1 Expert data

In order to estimate the parameters of the three stochastic processes to the expert data we used different values for α_1, α_2 and α_3, i.e. the probabilities assigned to each interval.

Process	μ (% year^{-1})	σ (% year^{-1})	q (-)
Table 2.3a	$\alpha_1 = \alpha_2 = 0.75$, $\alpha_3 = 0.8$		
BM	1.28E−3	7.40E−3	3.56
GP	1.34E−3	7.37E−3	3.55
Table 2.3b	$\alpha_1 = \alpha_2 = \alpha_3 = 0.8$		
BM	1.42E−3	5.38E−3	3.52
GP	1.42E−3	5.51E−3	3.52
Table 2.3c	$\alpha_1 = \alpha_2 = 0.8$, $\alpha_3 = 0.95$		
BM	9.80E−4	6.93E−3	3.67
GP	1.01E−3	7.10E−3	3.66
Table 2.3d	$\alpha_1 = \alpha_2 = 0.9$, $\alpha_3 = 0.95$		
BM	1.32E−3	3.89E−3	3.55
GP	1.32E−3	3.87E−3	3.55

Table 2.3: Parameter estimates for BM and the gamma process (expert data).

The parameter estimates and the sum of squared differences for both the gamma and the Brownian motion process are given in Table 2.3. The parameter estimates of these processes do not differ much; the relative difference between the corresponding parameters is highest for the drift parameter (2.8% in Table 2.3c). Note that the values of the power q are smaller than in the deterministic model. The relative difference between the volatility parameters (σ) is at most 2.3% (see Table 2.3c). In this case the variance of BM (gamma process), given by $\sigma^2 t^q$, grows from 0.22 (0.23) after 10 years to 2.9 (2.9) in year 20. This indicates that there is little uncertainty in the deterioration process, especially in the first 10 years. The small differences between the processes are explained as follows. Firstly, the large drift dominates possible negative increments of Brownian motion and secondly, the variance of the processes is very small.

On the other hand, the computing time of the parameter estimation procedures is not equal. Whereas evaluating the SSD for four billion combinations of parameter values takes about 7 hours for the BM process, it takes almost 13 hours to check the same

number of combinations for the gamma process. The main reason for this difference in computing time is that the evaluation of the gamma functions in Equation (2.5) is more time-consuming than the evaluation of the cumulative standard normal distribution functions in Equation (2.2).

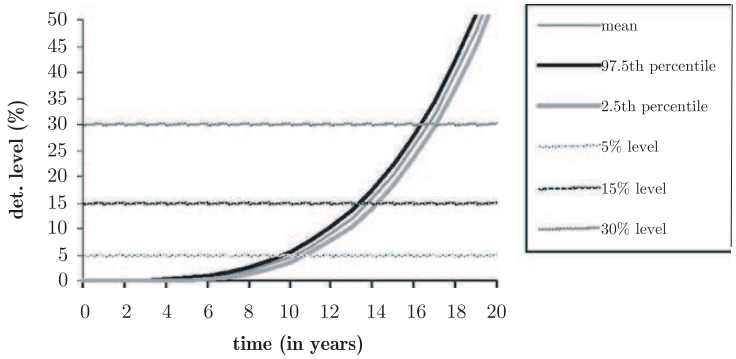

Figure 2.1: Mean and 2.5th and 97.5th percentiles of Brownian motion.

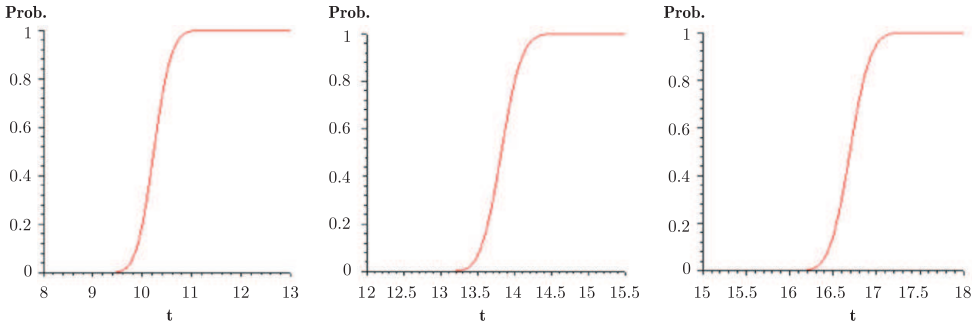

Figure 2.2: Cumulative distribution functions of the hitting time of Brownian motion at 5%, 15% and 30%.

Figures 2.1 and 2.2 are based on the parameter estimates given in Table 2.3c. In Figure 2.1 the expected course of BM and the 2.5th and 97.5th percentiles are plotted against time. The graphs show that the deterioration is between 18% and 22% after 15 years, with probability 0.95. Drawing these graphs for the gamma process gives similar results. In Figure 2.2, cumulative distribution functions of the hitting time of BM are plotted for the levels 5%, 15% and 30%. The graph at the left hand side shows that the probability that the deterioration level exceeds 5% in the first 10.5 years is 0.83. Thus, the middle of the interval is not a good point estimate for the time that this level is exceeded. Note that the probability that the deterioration level exceeds 30% before year 16 is 0. This shows that the experts have mentioned a too wide interval in comparison

to the other two intervals; an interval $[16, 17]$ would be better, since the probability that the deterioration level exceeds 30% in this interval is 0.95.

The TSHG process has been fitted to the expert data for two different sets of interval probabilities: $[0.8, 0.8, 0.95]$ and $[0.75, 0.75, 0.8]$. The parameter estimates can be found in Tables 2.4 and 2.5. The computing time of the estimation procedure is much higher than it is for the gamma process and BM models. It takes about 10 hours to evaluate the stochastic objective function (based on 10 000 simulation runs) for (only) 10 000 different parameter combinations. The estimates of p in the TSHG processes with non-stationary initiation phase, indicate that the initiation of spots is indeed non-stationary. In Table 2.6 (Table 2.7) it is seen that the TSHG process with non-stationary initiation phase and deterministic interexpansion time fits the interval probabilities $[0.8, 0.8, 0.95]$ ($[0.75, 0.75, 0.8]$) best. In this case the value of the objective function for this parameter vector, based on 50 000 simulations, is equal to 1.73E−3 (2.16E−4).

Tables 2.4 and 2.5 also show that the estimated fractions $G_i(\theta)$ are not as close to the probabilities α_i as we expected them to be. The differences between $G_i(\theta)$ and α_i can be explained by the relatively small number of parameter combinations for which the stochastic objective function has been evaluated.

Initiation	Propagation	λ (% year^{-1})	p (-)	δ (year)
Stationary	Deterministic	2.14	0	1.89
Non-stationary	Deterministic	0.16	2.11	2.57
Non-stationary	Stochastic	0.028	2.85	3.55

Table 2.4: Parameter estimates for 3 versions of the simulated TSHG process for $\alpha_1 = \alpha_2 = 0.8$ and $\alpha_3 = 0.95$ (expert data).

Initiation	Propagation	$G_1(\theta)$	$G_2(\theta)$	$G_3(\theta)$	$f(\theta)$
Stationary	Deterministic	0.114	0.312	0.474	0.94
Non-stationary	Deterministic	0.768	0.774	0.953	1.73E−3
Non-stationary	Stochastic	0.628	0.695	0.897	4.35E−2

Table 2.5: Goodness-of-fit measures for 3 versions of the simulated TSHG process for $\alpha_1 = \alpha_2 = 0.8$ and $\alpha_3 = 0.95$ (expert data).

Figures 2.3-2.5 are based on (simulations of) the TSHG process with the parameter estimates given in the third row of Table 2.4. It appears from Figure 2.3 that the 2.5th and 97.5th percentiles of the two-stage hit-and-grow model result in relatively wide 95% probability intervals, as compared to the other two processes. This indicates that the

Initiation	Propagation	λ (% year^{-1})	p (-)	δ (year)
Stationary	Deterministic	2.35	0	2.02
Non-stationary	Deterministic	0.058	2.65	2.55
Non-stationary	Stochastic	0.025	2.90	3.53

Table 2.6: Parameter estimates for 3 versions of the simulated TSHG process for $\alpha_1 = \alpha_2 = 0.75$ and $\alpha_3 = 0.8$ (expert data).

Initiation	Propagation	$G_1(\theta)$	$G_2(\theta)$	$G_3(\theta)$	$f(\theta)$
Stationary	Deterministic	0.115	0.301	0.443	0.733
Non-stationary	Deterministic	0.751	0.750	0.785	2.16E−4
Non-stationary	Stochastic	0.637	0.681	0.859	2.09E−2

Table 2.7: Goodness-of-fit measures for 3 versions of the simulated TSHG process for $\alpha_1 = \alpha_2 = 0.75$ and $\alpha_3 = 0.8$ (expert data).

two-stage hit-and-grow model predicts a more volatile deterioration process. Figure 2.4 shows a realization of this model. It is seen that the percentage of surface to be repaired, computed as the area of the black spots, is 27.71% after 15 years. Apart from knowing the percentage of surface to be repaired, the distribution of the spots on the surface also provides valuable information. In Figure 2.5 a histogram of the size of the spots on the surface shown in Figure 2.4 is given. It follows that there are many spots of size 1, 9 and 25 (measures in unit squares); this is because infected squares infect all their neighbours. Secondly, in this case the total area covered by spots of size less than or equal to (greater than) 25 is 718 (2053) unit squares. So, it is relatively easy to paint about 75% of the deteriorated surface in this case; the other 25% consists of small spots.

2.6.2 Inspection data

Estimation results Brownian motion and gamma process

In Table 2.8 the maximum likelihood estimates of the parameters of Brownian motion and the gamma process are given. The difference between the values of the powers q is striking. To compensate for the high value of q, the parameter values of μ and σ are relatively small for Brownian motion. It appears that for the first 20 years the variance (given by $\sigma^2 t^q$) of the gamma process is larger than the variance of BM. Comparing the quality of the fit of the two processes is difficult in this case, since we only have 5 data points. The fact that the log likelihood value is higher for BM does not automatically imply that BM fits the data better than the gamma process. After all, the two models are

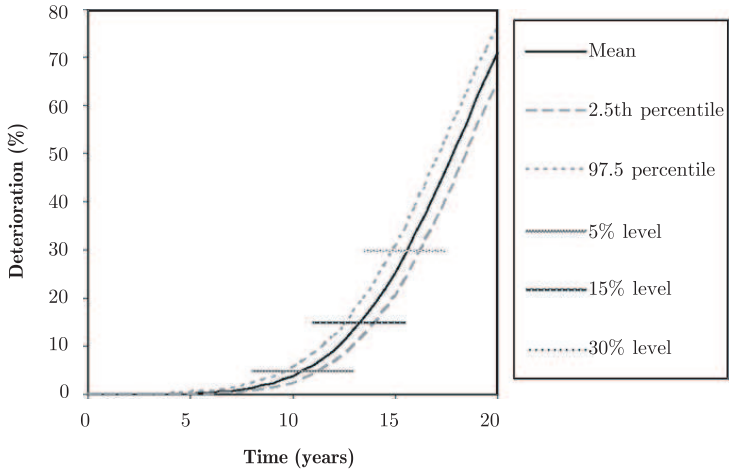

Figure 2.3: Expected cumulative deterioration for the two-stage-hit-and-grow process with non-stationary arrival process and deterministic interexpansion times.

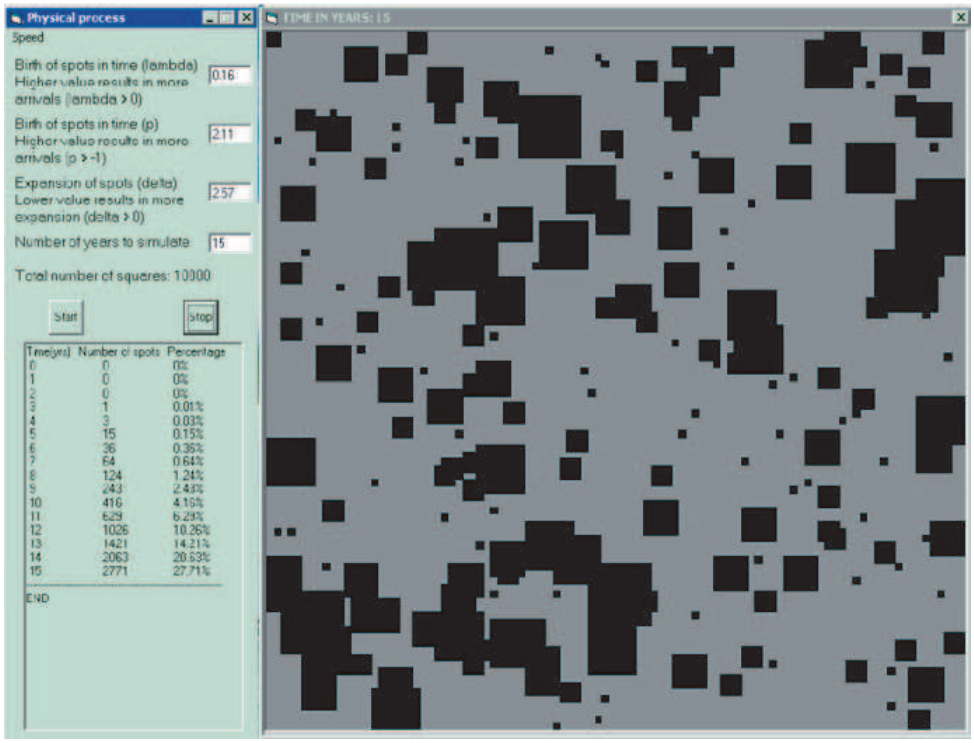

Figure 2.4: Realization of the TSHG process with non-stationary arrival process and deterministic interexpansion times.

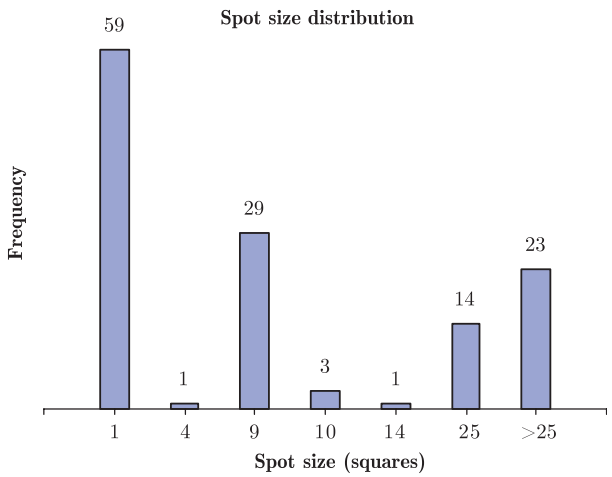

Figure 2.5: Spot size distribution of the realization of the TSHG process shown in Figure 2.4.

non-nested. It appears that the estimation of both processes is equally time-consuming (a few seconds).

Parameter	Brownian motion	Gamma process
μ	1.76E−3	3.29E−3
σ	9.56E−3	1.42E−2
q	2.63	2.37
Log likelihood	1.31	0.825

Table 2.8: Parameter estimates for Brownian motion and the gamma process (inspection data).

Figures 2.6 and 2.7 show the expected condition with its 5th and 95th percentiles including the inspection data, the probability density function of the lifetime (time to failure), and the cumulative distribution function of the lifetime according to the gamma process and Brownian motion, respectively. The deterioration failure level is a corroded surface of 3%. The lifetime is the time in years until the failure level is exceeded. Looking at the expected condition and the 5th and 95th percentiles according to BM and the gamma process, both processes fit the data well. According to the BM the lifetime is between 15.6 and 18.4 years with probability 0.95; according to the gamma process it is between 16.0 and 19.5 years with probability 0.9. The mean lifetimes are given by 16.9 years for BM and 17.7 years for the gamma process. As opposed to the gamma process, BM can result in a negative deterioration. This can be seen by comparing the 5th per-

centile of the deterioration for both processes in Figure 2.6 (top) and Figure 2.7 (top). For BM, the 5th percentile of the deterioration can be even negative! The larger the uncertainty in the deterioration process, the higher the probability of negative deterioration for BM, and the less suitable BM is for modelling deterioration.

Estimation results TSHG process

As the TSHG process measures the percentage of the area to be repaired, we multiply the values in Table 2.2 by a factor 7 in order to estimate the parameters of the process. The (log) likelihood function corresponding to the TSHG process was approximated by doing 50 000 simulations. This makes the estimation a time-consuming process in comparison with the BM and GP models. The parameter estimates are found in Table 2.9. The sign of p is conspicuous. A negative value of this parameter implies that the intensity rate function decreases with time (the mean value function is increasing, but at a decreasing rate). This contradicts some intuition on this parameter, but it may reflect mediocre/bad application of the paint. The value of parameter δ implies that spots expand every 2.5 years.

Parameter	TSHG process
λ	2.81
δ	2.53
p	-0.42
Log likelihood	1.77

Table 2.9: Parameter estimates for the TSHG process (inspection data).

In Figure 2.8 the expected cumulative deterioration including the inspection data and 5th and 95th percentiles are shown. The expected cumulative deterioration and the 5th and 95th percentiles are computed by performing 500 000 simulations. Looking at the expected condition and the 5th and 95th percentiles the TSHG process fits the data well. The variance of the TSHG process appears to be larger than the variance of BM and the gamma process.

In Figure 2.9 the empirical cumulative distribution function of the lifetime according to the TSHG process are shown. In order to construct the empirical CDF of the lifetime 100 000 realizations of the hitting time at 21% have been simulated with the TSHG process. Note that the deterioration failure level is a corroded surface of 3%, which corresponds with an area to be repaired of 21%. The lifetime is the time in years until the failure level is exceeded. According to this model the lifetime is between 15.1 and 20.9 years with probability 90%. This is a much broader interval than we obtained for

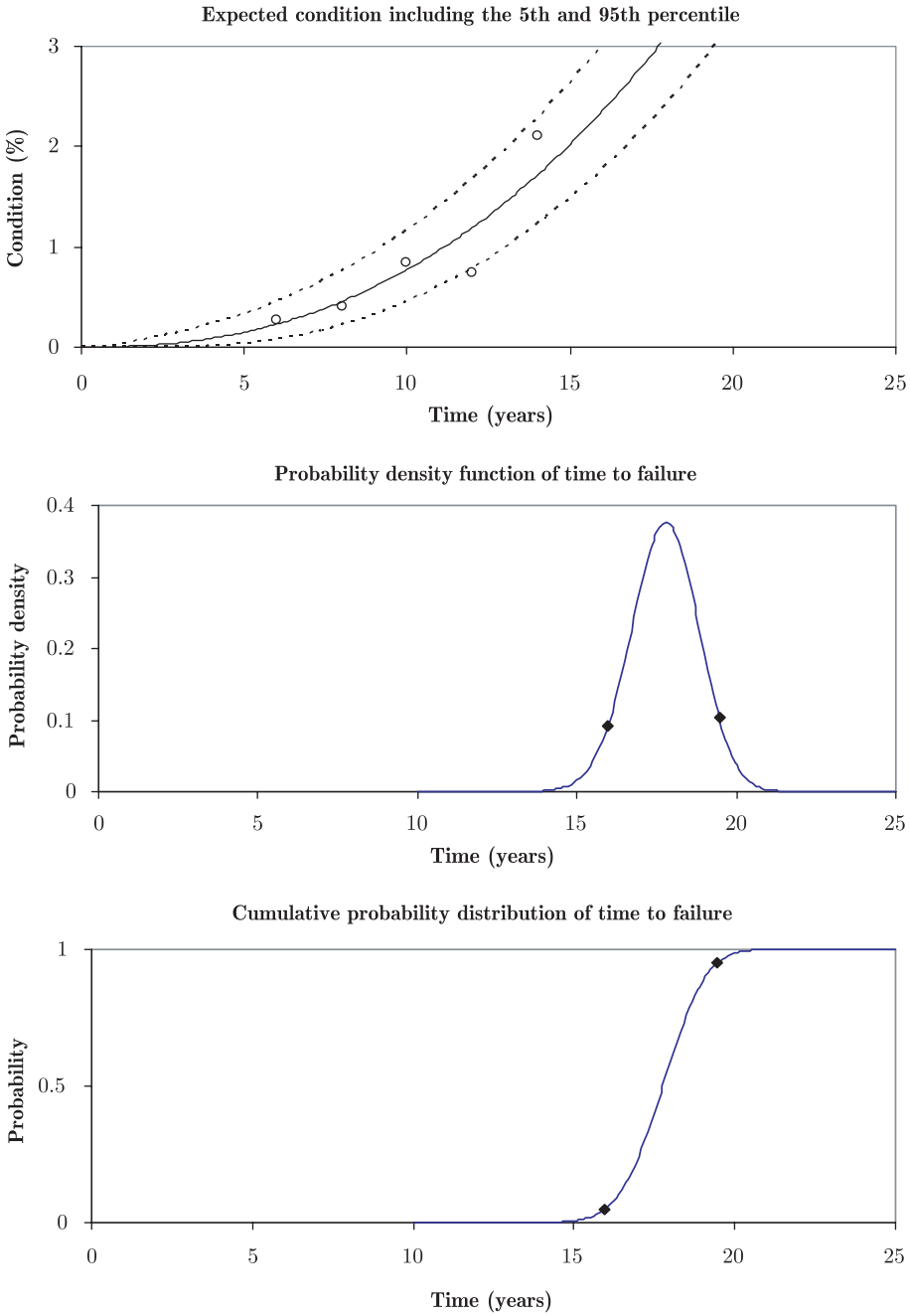

Figure 2.6: Expected cumulative deterioration, based on the maximum likelihood estimates of the gamma process, including five inspection data (○) and the 5th and 95th percentiles (—); Probability density function of the lifetime and its 5th and 95th percentiles (⧫); Cumulative distribution function of the lifetime and its 5th and 95th percentiles (⧫).

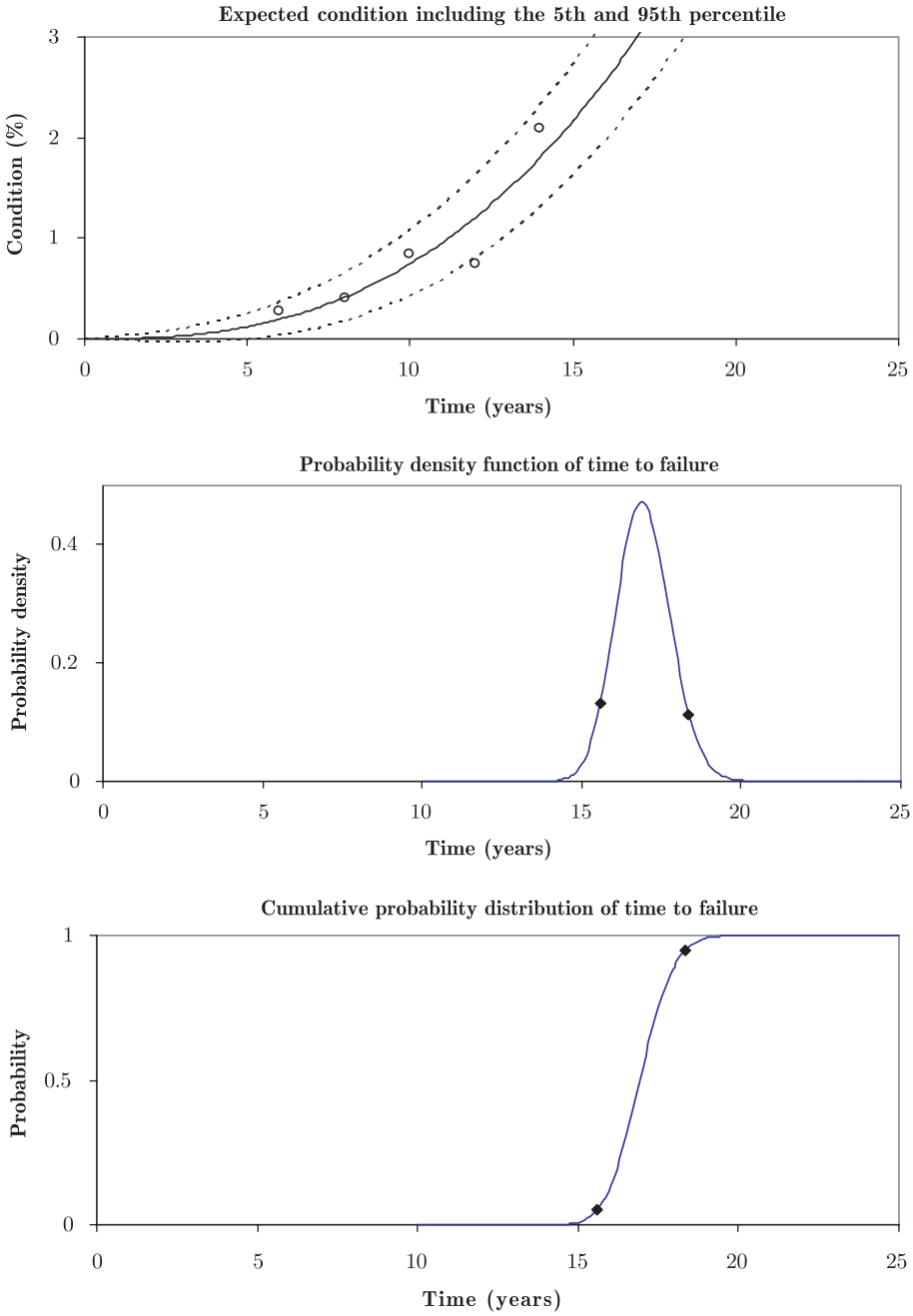

Figure 2.7: Expected cumulative deterioration, based on the maximum likelihood estimates of Brownian motion, including five inspection data (∘) and the 5th and 95th percentiles (—); Probability density function of the lifetime and its 5th and 95th percentiles (⧫); Cumulative distribution function of the lifetime and its 5th and 95th percentiles (⧫).

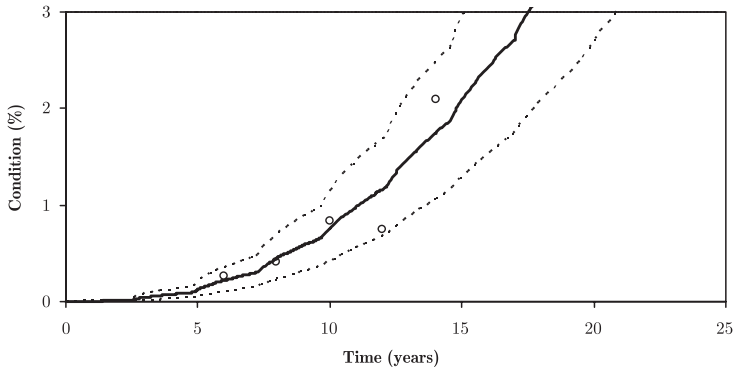

Figure 2.8: Expected cumulative deterioration, based on the maximum likelihood estimates of the TSHG process, including five inspection data (∘) and the 5th and 95th percentiles (—).

BM and the gamma process. The mean lifetime is 17.8 years, which corresponds with the result obtained by the gamma process.

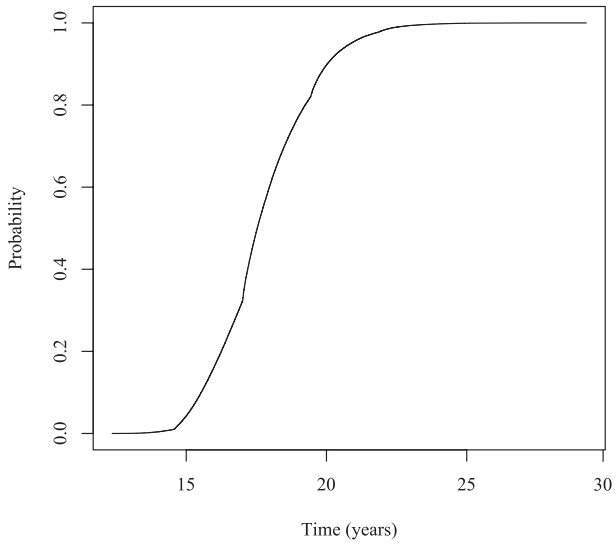

Figure 2.9: Cumulative distribution function of the lifetime, constructed from 100 000 simulated hitting times from the TSHG process.

2.6.3 Meta-analysis

In order to determine whether the TSHG process can be replaced by a BM or a gamma process, we have fitted these processes to the interval probabilities resulting from the

TSHG process. By simulation we have first determined the probability that the TSHG process, having deterministic interexpansion times, exceeds 7 given levels in 7 given intervals for the parameter vectors $\theta = (0.16, 2.11, 2.57)$ and $\theta = (0.058, 2.65, 2.55)$. The levels, intervals and the resulting probabilities (and their standard errors) are given in Table 2.10. The results are obtained by doing 250 000 and 1 000 000 simulations respectively, ensuring that the variance of the estimators is at most 1E−6.

		$\theta = (0.16, 2.11, 2.57)$	$\theta = (0.058, 2.65, 2.55)$
Level %	Interval (years)	prob. (std. error)	prob. (std. error)
1	$7 - 8$	0.682 (9.3E−4)	0.500 (4.4E−4)
2	$8 - 9$	0.661 (9.5E−4)	0.432 (3.5E−4)
5	$10 - 11$	0.777 (8.3E−4)	0.752 (4.5E−4)
10	$12 - 13$	0.662 (9.5E−4)	0.744 (4.6E−4)
15	$13 - 14$	0.750 (8.7E−4)	0.750 (4.8E−4)
20	$14 - 15$	0.632 (9.6E−4)	0.481 (3.9E−4)
30	$15 - 17$	0.935 (4.9E−4)	0.785 (3.1E−5)

Table 2.10: Estimated interval probabilities and their standard errors.

Next, the parameters of Brownian motion and the gamma process are estimated in the way explained in Section 2.5.1. The estimates and the sum of squared differences are given in Table 2.11. If we compare these estimates with the parameter estimates of Brownian motion and gamma process fitted directly to the probabilities, we notice that the resulting processes are quite different. For example, the variance term is twice as small, whereas the power has increased by more than 25%.

Parameter	$\theta = (0.16, 2.11, 2.57)$		$\theta = (0.058, 2.65, 2.55)$	
Process	BM	GP	BM	GP
μ	7.01E−5	9.78E−5	2.12E−5	3.02E−5
σ	3.22E−3	3.75E−3	2.20E−3	2.39E−3
q	4.75	4.63	5.21	5.08
SSD	1.38E−2	7.89E−3	1.53E−2	6.08E−3

Table 2.11: Parameter estimates and sum of squared differences (SSD) for BM and the gamma process fitted to the simulated data.

The results in Table 2.12 show that the absolute differences between the interval probabilities of the hit-and-grow model on the one hand and the probabilities according to BM and the gamma process on the other hand are substantial. Thus, the grey-box processes are no perfect substitutes for the TSHG process.

Parameter		$\theta = (0.16, 2.11, 2.57)$		$\theta = (0.058, 2.65, 2.55)$	
Level %	Interval (years)	BM	GP	BM	GP
1	$7-8$	0.088	0.068	0.020	0.020
2	$8-9$	0.028	0.021	0.080	0.053
5	$10-11$	0.038	0.032	0.052	0.015
10	$12-13$	0.017	0.011	0.002	0.002
15	$13-14$	0.052	0.039	0.036	0.032
20	$14-15$	0.030	0.017	0.040	0.022
30	$15-17$	0.009	0.004	0.054	0.034

Table 2.12: Absolute differences between the (estimated) interval probabilities of Brownian motion and the gamma process, and the (simulated) probabilities for the TSHG process.

2.7 Conclusions

We have compared three stochastic processes that can be used to model the deterioration of organic coating systems protecting steel structures. We have fitted these processes to expert data (for a specific group of steel objects in the Netherlands) and to inspection data (for the Haringvliet steel gates). In both cases it is found that the deterioration of organic coatings is non-stationary. The expected deterioration follows a power law function with power greater than 1, meaning that the coatings deteriorate faster and faster.

Brownian motion with a non-linear time transformation and the gamma process with non-linear shape function describe the non-stationary deterioration process very well. Both processes arise from the standard form of the stochastic process by applying a time transformation. It appears that the differences between these processes are small. The reason is that the drift of the processes is high and the variance is relatively small. In spite of this, Brownian motion may have negative increments. On the other hand, the estimation of the gamma process from expert data is more time-consuming.

We have also introduced a two-stage hit-and-grow process, which is founded on the physical properties of the deterioration of coatings. The simulation of this process is time-consuming but necessary, since it is difficult to determine its probabilistic properties analytically. We have fitted different versions of this process to the expert data. It appears that some of these versions fit the data well. For example, choosing the initiation process to be a Poisson process with non-linear expectation gives a good fit. Also, the version with constant (deterministic) interexpansion times performs better than the version with stochastic interexpansion times. All by all, the two-stage hit-and-grow process gives more insight into the physics of the actual deterioration process than the grey-box models. For

example, by using the TSHG model the initiation rate and the time a corrosive spot needs to expand can be assessed. This should not surprise since the process relates more to the physics of the deterioration of coatings. We have also fitted the TSHG process to the inspection data for the Haringvliet steel gates. To this end, the likelihood of the TSHG process has been estimated by applying kernel density estimation to the simulation output. It appears that the process fits the data well. Surprisingly, the initiation rate of the spots decreases with time.

A disadvantage of the TSHG process is that the probabilistic characteristics have to be determined (approximated) by means of simulation. As a consequence computing the probability distributions of the TSHG process and its hitting time is time-consuming. Therefore it seems better not to use the simulation model in a decision support system for maintenance optimization. A meta-analysis was performed in order to see whether Brownian motion or the gamma process can replace the TSHG process in such a system. We found that fitting Brownian motion and the gamma process to interval probabilities of the TSHG process may result in different results as compared to the direct fitting of these two grey-box processes.

Possible extensions of this work are the introduction of imperfect maintenance actions such as spot repair and repainting in the deterioration processes. The TSHG process is especially suited for this, since we can define different arrival processes for painted spots and unrepaired spots. The gamma process and Brownian motion models are also to be modified in order to take into account maintenance actions. A related extension is modelling the maintenance optimization problem for steel structures protected by coating systems. Deterioration modelling plays a big role in this problem. In Chapter 6 we will come back to these extensions.

Chapter 3

A framework for statistical inference on discrete event systems*

Abstract

In this chapter we present a framework for statistical analysis of discrete event systems. This framework combines tools such as simulation of marked point processes, likelihood methods, kernel density estimation and stochastic approximation, and is applicable even if conventional approaches fail due to the mathematical intractability of the discrete event system. The approach is illustrated with an application to modelling and estimating corrosion of steel gates in the Dutch Haringvliet storm surge barrier.

3.1 Introduction

Uncertainty is an important part of life, and hence statistical modelling is appropriate for many real life phenomena. Unfortunately, the complexity of statistical models is a trade-off between the mathematical tractability of the model on one hand, and insight into the structure of the phenomenon of interest on the other hand. On the one side of the spectrum there are black-box models, which are easily analyzed statistically, but do not provide any deep structural insights. On the other side of the spectrum there are white-box models, which carefully set out in detail structure, but are hardly amenable for statistical analysis.

We focus on discrete event systems, dynamic systems that result in the occurrence of events at particular points in time; each event has a certain discrete type. In the context of discrete event systems, modelling the time to occurrence of a certain event by choosing

*This chapter is based on Nicolai and Koning (2006)

a distribution from some catalogue of distributions (for instance, Weibull, gamma, log-normal) is an example of the black-box approach, whereas a full-fledged simulation model is an example of a white-box model.

Our primary motivation for undertaking this study comes from experience in the field of deterioration modelling, see Chapter 2 and Nicolai *et al.* (2007a). In this field, there is an abundance of grey-box models which are motivated by but do not fully capture the underlying physical process of deterioration. Examples are deterioration models derived from Brownian motion, gamma or compound Poisson processes, see Doksum and Hóyland (1992); Çinlar *et al.* (1977); Sobczyk (1987). These models have been successfully applied to quantify deterioration such as wear, corrosion, crack growth and creep.

As black-box models are available for purely descriptive purposes, the abundance of grey-box deterioration models underlines the need for incorporating structural insights into the statistical models. Nevertheless, white-box models, the pinnacles of structural thinking, are not used in the statistical analysis. The reason for this is that conventional approaches to statistical analysis are not feasible due to the mathematical intractability of the white-box models. For instance, applying likelihood methods is not feasible since computing the likelihood becomes too complex.

The objective of this study is to develop a likelihood based methodology for statistical analysis of discrete event systems by means of white-box models, in which the mathematical intractability of the white-box model is circumvented by using computer intensive methods. In particular, the likelihood is 'maximized' using simulation, density estimation and stochastic approximation. Thus, even if the likelihood of a white-box model cannot be computed numerically, the model can still be estimated. As such our methodology hands us a potentially powerful tool for statistical inference for white-box models. This is especially key for white-box deterioration models which should take into account physical relations.

Although the elements of our methodology already exist separately in literature, it is their combination which is new, and opens up the possibility of statistical inference for white-box models in new application areas.

As an illustration, we present a white-box model for steel corrosion and fit this to the Haringvliet corrosion data presented by Frangopol *et al.* (2004). The Dutch Haringvliet storm surge barrier was built in 1970, and closes the sea arm Haringvliet off from the sea. In the period from 1988 to 1996, the original coatings were replaced by new ones. In 2002, half of these coatings were inspected resulting in different inspection intervals. In Table 3.1 the percentage of the seaside surface of a steel gate that has been corroded due to ageing of the coating is given for five steel gates of the Haringvliet barrier.

In Frangopol *et al.* (2004) probabilistic models for deteriorating structures, such as the steel gates of the Haringvliet barrier, are reviewed. Although many models are discussed,

Age (years)	Corrosion (%)
6	0.27
8	0.41
10	0.84
12	0.75
14	2.10

Table 3.1: Inspection results for the seaside surface of five steel gates of the Haringvliet storm surge barrier.

none of them is white-box. The methodology in this study does allow us to estimate a white-box model for the Haringvliet corrosion data.

To model the corrosion process of a steel gate as a discrete event system, we partition the rectangular surface of the steel gate in small square areas, which we will refer to as sites. Each of these sites can be in two states; initially, the square is 'uncorroded', but after some time it may become corroded. Once the site is corroded, it cannot return to the uncorroded state.

This chapter is organized as follows. In Section 3.2 we introduce our methodology, in Section 3.3 we illustrate the approach by analyzing the Haringvliet corrosion data with a white-box model, and in Section 3.4 we draw conclusions.

3.2 Methods

3.2.1 Model building

Capturing the white-box model

In this section we present a methodology for statistical analysis of discrete event systems given a white-box model. As the methodology is likelihood based, we first need to capture the white box model in mathematical equations, which allows us to identify the parameters to be estimated.

We achieve this by formulating the discrete event system in terms of a marked point process, a stochastic process which consists of events of a certain type (the mark) taking place at certain points of time. To each marked point process belongs a counting process $N^* = \{N^*(t) : \ t \geq 0\}$, which counts the number of occurrences of events in the interval $[0, t]$. Note that the counting process increases by 1 at every occurrence of any event. Thus, every 'jump time' of the counting process coincides with an occurrence of some event, and vice versa. In addition, to each marked point process there belongs a mark sequence $S_1, S_2, \ldots$, which accommodates the event type: the i^{th} mark S_i gives the type

of the event occurring at the i^{th} jump time. The marks S_i take values in some mark space $\mathcal{S}$. For discrete event systems, the mark space $\mathcal{S}$ is discrete. For instance, the corrosion of steel gates may be described by a marked point process which records every point of time at which a site becomes corroded, and assigns to this time the location of the site as a mark. Thus, in this particular case the mark space $\mathcal{S}$ is in fact the product of a finite 'horizontal' space $\mathcal{S}_x$ and a finite 'vertical' space $\mathcal{S}_y$.

Let us return to the general situation. As for every possible mark value $s \in \mathcal{S}$ there exists a counting process $N_s = \{N_s(t) : \ t \geq 0\}$, which counts the number of occurrences of events of type s in the interval $[0, t]$, we may alternatively view the marked point process as a multivariate counting process $\{N_s\}_{s\in\mathcal{S}}$. Moreover, we may write $N^* = \sum_{s\in\mathcal{S}} N_s$.

As a counting process is a submartingale, it follows as a result of the Doob-Meyer decomposition that there exists a corresponding compensator, that is, a nondecreasing predictable process such that the difference of the counting process and its compensator is a martingale, see (Andersen *et al.*, 1993, p. 73). Predictability may be characterized as follows: the value of a predictable process at time t is completely determined just before t, see (Andersen *et al.*, 1993, p. 67).

The derivative of the compensator with respect to time is called the intensity rate function. Let $\Lambda^* = \{\Lambda^*(t) : \ t \geq 0\}$ and $\Lambda_s = \{\Lambda_s(t) : \ t \geq 0\}$ respectively denote the compensators, and let $\lambda^* = \{\lambda^*(t) : \ t \geq 0\}$ and $\lambda_s = \{\lambda_s(t) : \ t \geq 0\}$ respectively denote the intensity rate function belonging to N^* and N_s. We may write

$$\Lambda^*(t) = \int_0^t \lambda^*(s)ds, \quad \Lambda_s(t) = \int_0^t \lambda_s(s)ds.$$

It is known that the structure of the compensator completely describes the probabilistic behaviour of the corresponding counting process, and thus we may model a counting process by imposing structure on its compensator, or equivalently, imposing structure on its intensity rate function. As

$$\lambda^* = \sum_{s\in\mathcal{S}} \lambda_s, \tag{3.1}$$

it follows that a discrete event system may be modelled by imposing structure on each of the mark-specific intensity rate functions λ_s. We shall assume that $\lambda_s(t)$ takes the form $\lambda(t; \theta_s, \mathbf{z}_s)$, where θ_s is a vector of structural parameters, and $\mathbf{z}_s = \{\mathbf{z}_s(t) : \ t \geq 0\}$ is a vector of possibly time-dependent covariates. The covariate vector $\mathbf{z}_s$ contains all environmental factors determining the 'risk' of an event of type s. With regard to steel structures, the local weather conditions and the type of coating system applied to the structure are the environmental factors influencing corrosion most (see Nicolai *et al.*, 2007a). Note that λ_s should be predictable, that is, $\lambda_s(t)$ is completely determined just before t; hence, we will tacitly assume that a time-dependent covariate $\mathbf{z}_s$ only has effect on $\lambda_s(t)$ through its 'historical part' $\{\mathbf{z}_s(t')\}_{0<t'<t}$.

The covariate vector may include global as well as local environmental factors. An often useful way to handle local environmental factors is to incorporate the state of a neighbourhood system as covariate in the model. A neighbourhood system defines a neighbourhood $\mathcal{B}_s$ for each $s \in \mathcal{S}$. For example, if $\mathcal{S}$ is a two-dimensional grid (representing a surface) then the neighbourhood of s may contain the four sites directly north, west, south and east of s, see Figure 3.1. Note that s itself is not an element of $\mathcal{B}_s$.

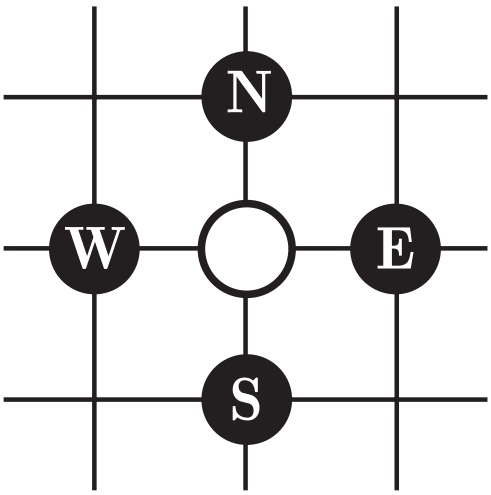

Figure 3.1: Example of a neighbourhood in a two-dimensional grid. The black dots N, W, S and E together form the neighbourhood of the white dot.

In model building, one typically starts with a baseline model which does not include any covariate. Then, the baseline model is extended by including covariates. This can be conveniently done by multiplying a baseline parameter by $r(\beta^T \mathbf{z}_s)$, where β is a vector of unknown regression coefficients, and r is a known scalar function. Typical choices are $r(v) = e^v$ and $r(v) = v$, yielding multiplicative and additive covariate effects, respectively.

Finally, we present a generic scheme for simulating a marked point process. It extends an existing algorithm for simulating counting processes (Daley and Vere-Jones, 2003, Algorithm 7.4.III, p. 260), by including the assignment of event types. The scheme involves $p_s(t)$, the probability that the next event of N^* is of type s given that it occurs at time t. Observe that $p_s(t) = \lambda(t; \theta_s, \mathbf{z}_s)/\lambda^*(t)$.

Algorithm 3.1 (Simulation scheme for marked point process $\{N_s\}_{s\in\mathcal{S}}$)
Initially, let t_0 be the starting time of the simulation.

(1) Draw E, an exponential random variable with mean 1.

(2) The next event time t_1 is defined as $t_1 = \{t \geq t_0 : \ \Lambda^*(t) = \Lambda^*(t_0) + E\}$.

(3) Select mark s with probability $p_s(t_1)$ by the inverse transform method.

(4) Set $t_0 = t_1$, update Λ^*, and return to step (1).

We generate marks by the inverse transform method in step (3) to allow for the use of common random numbers (CRNs), which will become relevant later. Other ways of generating random variables, such as acceptance-rejection, typically do not allow for the use of CRNs. Furthermore, fast methods relying on storing tables a priori, such as the alias method, are undesirable because the probabilities under consideration are time-dependent.

The observation scheme

Algorithm 3.1 allows us to provide a model for a phenomenon under study. The next step involves fitting this model to sampled data, observations obtained from the phenomenon. As it is usually not feasible to fully observe all aspects of the phenomenon, we have to resort to partial observation in practice: we focus on one or more particular quantitative aspects of the phenomenon, and measure them at one or more points in time. Thus, on top of the phenomenon there is some observation scheme which ultimately produces the sampled data, a finite sequence of measurements, say $(X_1, X_2, \ldots, X_n)$. In general, the random variables $X_1, X_2, \ldots, X_n$, will be neither independent nor identically distributed. Thus, we should view $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ as a random vector drawn from some multivariate distribution.

In order to be able to simulate from the multivariate distribution of $\mathbf{X}$, we should mimic the observation scheme in simulations by some n-dimensional function $\mathbf{w}$ which assigns to each marked point process $\{N_s\}_{s\in\mathcal{S}}$ simulated by Algorithm 3.1 a random vector $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ of measurements $w_1\left(\{N_s\}_{s\in\mathcal{S}}\right)$, $w_2\left(\{N_s\}_{s\in\mathcal{S}}\right)$, $\ldots$, $w_n\left(\{N_s\}_{s\in\mathcal{S}}\right)$.

Let us illustrate the relationship between $\mathbf{X}$ and $\mathbf{w}$ by means of an application to a single steel gate, which is inspected once at age τ years. Assume the observed percentage of corrosion for this gate is given by $\mathbf{X} = (0.27)$ at age $\tau = 6$. The 'vector' function $\mathbf{w}$ maps the point process $\{N_s\}_{s\in\mathcal{S}}$ at the inspection time τ to a percentage of corrosion. That is, $\mathbf{w}(\{N_s\}_{s\in\mathcal{S}}) = 100N^*(\tau)/|\mathcal{S}|$, with $|\mathcal{S}|$ the total number of sites.

3.2.2 Statistical inference

Likelihood

Next, we take the multivariate distribution of the random vector $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ as a model for $\mathbf{X}$. As the random vector $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ is directly obtained from the marked point process $\{N_s\}_{s\in\mathcal{S}}$, it follows that the multivariate distribution of $\mathbf{w}\left(\{N_s\}_{s\in\mathcal{S}}\right)$ has the same parameters as $\{N_s\}_{s\in\mathcal{S}}$, say $\theta_1, \theta_2, \ldots, \theta_k$. Collect these parameters in the k-dimensional parameter vector θ. Note that θ contains all distinct elements of θ_s for every $s \in \mathcal{S}$. Let $\Theta \subset \mathbb{R}^k$ be the set of all possible values of θ.

Since we now have a statistical model for $\mathbf{X}$ which depends on a parameter vector θ, we are ready for parametric statistical inference on θ. We will limit ourselves to likelihood methods. Although there are many approaches to statistical inference, likelihood methods have always been popular following their proposal more than eighty years ago, and consequently have become well understood.

The likelihood of a statistical model evaluated in an arbitrary element ϑ of parameter space Θ is given by $L(\vartheta) = f_{\mathbf{X}}(\mathbf{X}; \vartheta)$, where $f_{\mathbf{X}}(\cdot; \vartheta)$ is the joint density of $\mathbf{X}$ under this model. In simpler models we are able to derive a closed mathematical expression for $L(\vartheta)$. Unfortunately, as we are dealing with white-box models, expressing $L(\vartheta)$ in this way is not feasible due to the mathematical intractability of the white-box models.

However, we do have a simulation model, which allows us to generate data from the joint distribution of $\mathbf{X}$ under the statistical model. As the likelihood $L(\vartheta)$ is in essence a joint density evaluated in $\mathbf{X}$, we may resort to density estimation based on the simulated data to obtain an approximation to $L(\vartheta)$. Therefore, we independently run the simulation m times for the parameter value ϑ. As the j^{th} simulation run yields a n-dimensional simulated vector of observations $\mathbf{X}^{(j)}$, we obtain independent random vectors $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(m)}$, all obeying the common unknown density $f_{\mathbf{X}}(\cdot; \vartheta)$.

For ease of exposition, we will in first instance focus on the univariate case $n = 1$, in which the sequence of random vectors $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(m)}$ coincides with a sequence of independent random variables $X^{(1)}, X^{(2)}, \ldots, X^{(m)}$. Define the kernel density estimator $\hat{f}(x)$ by

$$\hat{f}(x) = \frac{1}{mb} \sum_{j=1}^{m} K\left(\frac{X^{(j)} - x}{b}\right).$$

Here the kernel K is some probability density, symmetric around zero, and satisfying $\int (K(x))^2 \, dx < \infty$. The smoothing index b is often referred to as the bandwidth. Selecting the optimal bandwidth is an important topic in density estimation. A popular choice is the direct plug-in (DPI) bandwidth selector described in (Wand and Jones, 1995, paragraph 3.6.1). A variant of the DPI method is the 'solve the equation' plug-in bandwidth proposed in Sheather and Jones (1991). The Sheather-Jones plug-in bandwidth has been implemented in most statistical packages, and is widely recommended due to its overall good performance, see Sheather (2004).

Often, univariate density estimation (UDE) will suffice as we may partition $\mathbf{X}$ into n independent elements. For example, let us consider the data in Table 3.1; $\mathbf{X} = (0.27, 0.41, 0.84, 0.75, 2.10)$. Following Frangopol *et al.* (2004) we assume the inspection results to be independent. Now the joint density of the random vector $\mathbf{X}$ is the product of the densities of its independent components. Evaluating the density in the inspection results yields

$$\hat{f}(\mathbf{X}) = \hat{f}_{\mathbf{X}_1}(0.27)\hat{f}_{\mathbf{X}_2}(0.41)\hat{f}_{\mathbf{X}_3}(0.84)\hat{f}_{\mathbf{X}_4}(0.75)\hat{f}_{\mathbf{X}_5}(2.10).$$

Although UDE generally suffices, there are situations where we can only partition $\mathbf{X}$ into n_d independent components of length d_i, where $d_i > 1$ and $\sum_{i=1}^{n_d} d_i = n$. This forces us to rely on multivariate density estimation, a straightforward extension of UDE. The general expression for the multivariate kernel estimator is

$$\hat{f}(\mathbf{x}) = \frac{1}{mb} \sum_{j=1}^{m} K\left(\frac{\mathbf{X}^{(j)} - \mathbf{x}}{b}\right),$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, b is the bandwidth, and the kernel K is some n-variate square integrable probability density, symmetric around the origin. Work on the selection of the optimal bandwidth in multivariate density estimation is currently in progress, see Duong and Hazelton (2005).

One of the problems in multivariate density estimation is the 'curse of dimensionality', see (Scott, 1992, Section 7.2). This relates to the fact that the convergence of any estimator to the true value of a smooth function defined on a space of high dimension is very slow. In fact, the number of simulations m required to attain a specified amount of accuracy grows at least exponentially as n increases. Thus, when partitioning the vector $\mathbf{X}$ into independent components, one should preferably make the d_i's as small as possible. One may also appeal to data reduction techniques such as principal components or projection pursuit, see (Scott, 1992, Section 7.2). Finally, one may take the 'curse of dimensionality' into account in the choice of the observation scheme.

Stochastic approximation

By evaluating the density estimator $\hat{f}(\mathbf{x})$ for $\mathbf{x} = \mathbf{X}$, we obtain an approximation

$$\hat{L}(\vartheta) = \hat{f}(\mathbf{X}) = \frac{1}{mb} \sum_{j=1}^{m} K\left(\frac{\mathbf{X}^{(j)} - \mathbf{X}}{b}\right).$$

for the likelihood function $L(\vartheta)$. We will refer to this approximation as the *simulated* likelihood. The method of maximum likelihood dictates that the parameters of the model should be estimated by maximizing $L(\vartheta)$. Unfortunately, we are unable to evaluate $L(\vartheta)$ (and its gradient), and we have to make use of $\hat{L}(\vartheta)$, a noisy version of $L(\vartheta)$.

This kind of optimization problem can be handled in various ways, which differ in the way the gradient of $L(\vartheta)$ is estimated. The estimated gradient can be used in a steepest ascent procedure, where in each step additional simulation runs are performed. This method is generally referred to as stochastic approximation (SA). We will focus on stochastic approximation where the gradient is either estimated by means of finite differences (FD) or simultaneous perturbation (SP). These techniques only make use of the input and output of the underlying model.

The stochastic approximation algorithm builds upon the iteration formula $\hat{\theta}^{[\ell+1]} = \hat{\theta}^{[\ell]} + a_\ell \hat{\mathbf{g}}_\ell(\hat{\theta}^{[\ell]})$, where $\hat{\theta}^{[\ell]} \in \mathbb{R}^k$ represents the estimate of θ after the ℓth iteration, $\hat{\mathbf{g}}_\ell(\hat{\theta}^{[\ell]})$

is the approximation of the gradient at $\hat{\theta}^{[\ell]}$ and $\{a_\ell\}$ is a sequence of positive coefficients such that $a_\ell \to 0$ and $\sum_{\ell=1}^{\infty} a_\ell = \infty$.

In the ℓth iteration of the SA algorithm the two-sided FD approximation of the ith component of the gradient is given by

$$\hat{g}_{\ell i}(\hat{\theta}^{[\ell]}) = \frac{\hat{L}(\hat{\theta}^{[\ell]} + c_\ell e_i) - \hat{L}(\hat{\theta}^{[\ell]} - c_\ell e_i)}{2c_\ell}, \ i = 1, \ldots, k, \tag{3.2}$$

where e_i a unit vector in the ith direction and $\{c_\ell\}$ is a sequence of positive coefficients such that $c_\ell \to 0$ and $\sum_{\ell=1}^{\infty} a_\ell^2 c_\ell^{-2} < \infty$. Approximating the gradient in this way requires $2k$ evaluations of the simulated likelihood.

The SP method for estimating the gradient was first proposed in Spall (1987) and only requires a constant number of evaluations of the simulated likelihood. The idea is to perturb the parameters in separate random directions. In the two-sided SP approximation the ith component of the estimated gradient is given by

$$\hat{g}_{\ell i}(\hat{\theta}^{[\ell]}) = \frac{\hat{L}(\hat{\theta}^{[\ell]} + c_\ell \boldsymbol{\Delta}_\ell) - \hat{L}(\hat{\theta}^{[\ell]} - c_\ell \boldsymbol{\Delta}_\ell)}{2c_\ell \Delta_{\ell i}}, \ i = 1, \ldots, k, \tag{3.3}$$

where $\boldsymbol{\Delta}_\ell = (\Delta_{\ell 1}, \ldots, \Delta_{\ell k})^T$ is a vector of user-specified random variables satisfying certain conditions and c_ℓ is defined as before. Since the numerator in (3.3) is the same for each component i, only two function evaluations are required to obtain one SP gradient, that is, it does not depend on k.

Generally, one defines sequences $a_\ell = a\ell^{-\alpha}$ and $c_\ell = c\ell^{-\gamma}$, where $a, c, \gamma > 0$, $0 < \alpha \leq 1$. Now, under additional conditions on α, γ, $\boldsymbol{\Delta}_\ell$ (for the SP method), and the likelihood, $\hat{\theta}^{[\ell]}$ will converge almost surely to the 'true value' θ, for instance see Spall (2003). In Kleinman *et al.* (1999) it is shown that using the same sequence of random numbers for the simulation runs required to compute $\hat{L}(\hat{\theta}^{[\ell]} + c_\ell \boldsymbol{\Delta}_\ell)$ and $\hat{L}(\hat{\theta}^{[\ell]} - c_\ell \boldsymbol{\Delta}_\ell)$ has a positive effect on the rate of convergence of both the FD and the SP version of the SA algorithm (FDSA; SPSA). There is some evidence that the common random number version of SPSA should be preferred over its FDSA counterpart.

The scaled stochastic approximation algorithm described by Andradottir (1996) may be used to prevent the algorithm from diverging when the estimated gradient is either very steep or very flat. Finally, parameter transformations or projection based versions of the SA algorithm may be used to deal with parameter restrictions, for instance see Sadegh (1997).

3.3 Application to steel corrosion

3.3.1 A white-box model

In this section we will further specify the corrosion model for the data in Table 3.1. Recall that the counting process N_s records the time instance at which a site s becomes corroded. It follows from (3.1) that we may specify the model by formulating the structure of each site-specific intensity rate function λ_s belonging to N_s. As a site s cannot return to the uncorroded state once it is corroded, the corresponding counting process N_s can only jump once. In other words, the counting process is randomly stopped after the first jump. We may incorporate this into the model by including a factor $Y_s(t) = 1 - N_s(t-)$ in the intensity rate function λ_s. Here $N_s(t-)$ is shorthand notation for $\lim_{\epsilon \downarrow 0} N_s(t-\epsilon)$. In statistical modelling, the process Y_s is usually referred to as the 'number at risk', see (Andersen *et al.*, 1993, p. 128). Observe that the process Y_s is predictable. Next, we will assume that corrosion is the interplay of two physical processes: the initiation process and the propagation process.

In the initiation process the surface is constantly threatened by attacks. If an uncorroded site is hit, the site may or may not become corroded. However, as time progresses, the uncorroded sites become more and more vulnerable. We assume that the initiation process may be modelled by a (randomly stopped) inhomogeneous Poisson process with an intensity proportional to some power of t. That is, the initiation process contributes a term $\lambda_s^{[\text{initiation}]}(t) = q\nu t^{q-1} Y_s(t)$ to $\lambda_s(t)$.

In the propagation process, an uncorroded site s may become 'infected' by neighbouring corroded sites, that is, by corroded sites in its neighbourhood $\mathcal{B}_s$. We assume that $\mathcal{B}_s$ is as in Figure 3.1. The propagation process contributes a term $\lambda_s^{[\text{propagation}]}(t) = \delta z_s(t) Y_s(t)$ to $\lambda_s(t)$, where $z_s(t) = \sum_{s' \in \mathcal{B}_s} N_{s'}(t-)$ counts the number of corroded sites in $\mathcal{B}_s$ just before time t. It follows that

$$\lambda_s(t) = \lambda_s^{[\text{initiation}]}(t) + \lambda_s^{[\text{propagation}]}(t) = \left\{ q\nu t^{q-1} + \delta z_s(t) \right\} Y_s(t).$$

3.3.2 Results

In the previous subsection, we have formulated a model for the corrosion of a single steel gate. Assuming that the corrosion of one gate evolves independently of the corrosion of the other gates, this model is readily extended to the complete system of five steel gates. If in addition we assume that the five steel gates share the values of their parameters, we have $\theta = (\nu, q, \delta)^T$.

In order to approximate the likelihood function, the method of kernel density estimation is applied to the outcomes of $m = 1,000$ simulation runs of $\{N_s\}_{s \in \mathcal{S}}$, generated

according to Algorithm 3.1. Each outcome consists of the percentage of corrosion at the five inspection times.

We have applied the stochastic approximation algorithm introduced in section 3.2.2 to find the maximum likelihood estimates of the parameters of the corrosion process. In particular, we used the SPSA approach in combination with common random numbers (Kleinman *et al.*, 1999).

In Figure 3.2 the expected percentage of corroded area, the five inspection data, and 5th and 95th percentiles are plotted against time. The lifetime of a steel gate, given by the first passage time of the failure level (3% corrosion), has mean 17.8 years.

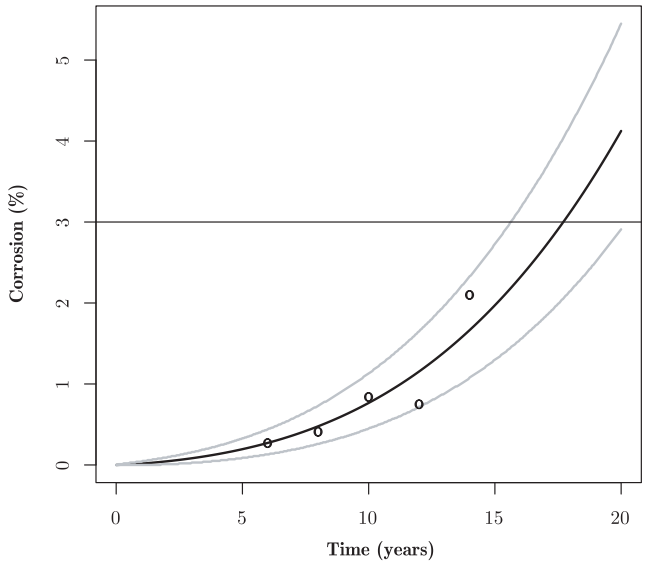

Figure 3.2: Expected percentage of corrosion (black line), 5th and 95th percentile (grey lines) including the five inspection data (∘) and the failure level (3% corrosion). The figure is based on $m = 50,000$ simulation runs of the white-box model with the estimated parameter vector $(\hat{\nu}, \hat{\delta}, \hat{q}) = (1.49\text{E}{-4}, 7.62\text{E}{-2}, 1.21)$. The estimated value of the likelihood is $\hat{L}(\hat{\nu}, \hat{\delta}, \hat{q}) = 5.35$. The expected lifetime is 17.8 years.

Figure 3.3 shows a realization of the corrosion process at time $t = 15$ years. The white pixels are uncorroded sites. The grey pixels are corroded sites due to initiation, and the black pixels are corroded sites due to propagation.

3.4 Conclusions

We have developed a methodology for statistical analysis of white-box simulation models by likelihood methods. The mathematical intractability of the white-box model is circum-

Figure 3.3: Simulated corrosion at $t = 15$ years.

vented by using computer intensive methods, including simulation, multivariate density estimation and stochastic approximation.

The objective in developing the methodology was complete generality. As a consequence, the methodology is indeed computer intensive, and may become very demanding on the available computing resources. For instance, due to the 'curse of dimensionality', multivariate density estimation may well become a time-consuming affair.

As reducing the computer intensiveness is a necessary condition for successful application to large problems, our methodology does not acquit the researcher from understanding the white-box model at a detailed level. For instance, a cleverly devised observation scheme is instrumental in averting the 'curse of dimensionality'.

Also, knowledge of the model may make solving $\Lambda^*(t) = \Lambda^*(t_0) + E$ in step (2) of Algorithm 3.1 simpler. For instance, in Section 3.3 we could have made use of the fact that corrosion is in fact the interplay of two physical processes.

Chapter 4

A comparison of two non-stationary deterioration models*

Abstract

Time-dependent deterioration of engineering structures can be modelled in several ways. Well-known models for this type of deterioration are the non-stationary gamma process and Brownian motion with a non-linear time transformation. Accordingly, the lifetime of a structure is modelled as the first hitting time of a fixed failure level by the stochastic process under consideration. In this chapter we compare the *implied* lifetime distributions for both processes under different deterioration scenarios. We also investigate the implications of the two stochastic processes to an age replacement policy for engineering structures in a simple maintenance model. In this model failures are *not* detected immediately, but only upon replacement (either preventive or corrective). A replacement restores the system to as new condition. We find some conspicuous differences between Brownian motion and the gamma process.

4.1 Introduction

Engineering structures such as tanks, bridges or pipelines are subject to time-dependent deterioration and appropriate maintenance actions should be done to protect them from failure. Failure of these structures is often considered normative and consequently the lifetime is the event in which deterioration exceeds a certain level, referred to as the failure level. Deterioration can be modelled by a stochastic process and in that case the lifetime is the first time the stochastic process under consideration exceeds the failure

*This chapter is based on Nicolai and Dekker (2007a)

level. As this lifetime is implied by the stochastic process it is also referred to as *implied* lifetime (Bae *et al.*, 2007).

A recent paper by Pandey *et al.* (2007) shows that the implications of the type of deterioration model to both the implied lifetime as well as the replacement policy of engineering structures can be considerable. The gamma process appears to be more versatile than a random variable model for stationary deterioration processes[1]. In particular, the gamma process takes into account temporal uncertainty and consequently it is a better model when the variability in deterioration, and thus the lifetime, is high. The question is whether this is also true when the deterioration process has a strong deterministic drift. We answer this question by comparing the non-stationary gamma process with another popular stochastic process in deterioration modelling, viz. Brownian motion with a non-linear time transformation. In particular we are interested in the first hitting times of these models. We also adopt the calibration scheme of Pandey *et al.* (2007) to compare the implications of both models on the lifetime and the optimal replacement age of engineering structures.

The literature on deterioration modelling is huge. We refer to Van Noortwijk (2007) for a literature review of the application of the gamma process in maintenance. Key articles on Brownian motion type deterioration models are Doksum and Hóyland (1992) and Whitmore and Schenkelberg (1997). Only few articles highlight the differences between these two models (see e.g. Van Noortwijk, 2007; Nicolai *et al.*, 2007a).

The chapter is organized as follows. In Section 4.2 we introduce a Brownian motion with a time transformation and the non-stationary gamma process. In Section 4.3 we give a thorough comparison of these stochastic processes. In particular, we focus on their first hitting times. In Section 4.4 we consider the role of both deterioration processes in a simple replacement model for engineering structures. In this model failure is normative and it is not detected immediately. Whereas Abdel-Hameed (1987) investigates a condition-based replacement policy for (a generalization of) this model, we optimize an age replacement policy. In Section 4.5 we draw conclusions.

4.2 Deterioration models

4.2.1 Brownian motion with a non-linear time transformation

We first consider Brownian motion with a time transformation. That is, let us define the amount of deterioration t time units after the last replacement of the component by

$$X(t) = \mu\Lambda(t) + \sigma W(\Lambda(t)),\ t > 0$$

[1] A stochastic process is called 'stationary' if the increments are independently distributed.

and $X(0) = 0$ w.p. 1. Here $\mu > 0$ is the drift parameter, $\sigma > 0$ is the volatility parameter, Λ is a function transforming time t, and W is a standard Brownian motion. In this study we assume Λ is a non-linear power function defined by $\Lambda(t) = t^q$, with $q > 0$ the power parameter. Consequently, the stochastic process $X = \{X(t),\ t \geq 0\}$ is a non-stationary Brownian motion with mean μt^q and variance $\sigma^2 t^q$. From the definition of Brownian motion it follows that the increments $X(t+h) - X(t)$ of the stochastic process X are independent and normally distributed with mean $\mu\left((t+h)^q - t^q\right)$ and variance $\sigma^2\left((t+h)^q - t^q\right)$, for all $t, h > 0$. Since the increments of the process can be both positive and negative, the process is not monotone. This has implications to the first hitting time of the failure level and accordingly to the age replacement policy under consideration; we will come back to this in Section 4.4.

The lifetime of the structure is defined as the first time the process X exceeds level ρ. If we denote this random quantity by T, then we may write $T = \inf\{t \geq 0 :\ X(t) > \rho\}$. Let the cumulative distribution function (CDF) of T be given by $F_{BM}(t)$, then it can be shown that (Karlin and Taylor, 1975)

$$F_{BM}(t) \;=\; \Phi\left(\frac{\mu t^q - \rho}{\sigma\sqrt{t^q}}\right) + \exp\left(\frac{2\mu\rho}{\sigma^2}\right)\Phi\left(\frac{-\mu t^q - \rho}{\sigma\sqrt{t^q}}\right), \tag{4.1}$$

where $\Phi(\cdot)$ denotes the CDF of a standard normal random variable. The first term in (4.1) represents the probability that X is above ρ at time t. The second term denotes the probability that the process exceeds ρ before t and it is below ρ at time t.

For a Brownian motion with linear drift ($q = 1$) the first hitting time of level ρ has an inverse Gaussian distribution with parameters ρ/μ and ρ^2/σ^2 (Doksum and Hóyland, 1992). In this special case, mean and variance of T are given by $\mathbb{E}(T) = \rho/\mu$ and $\mathrm{Var}(T) = \rho\sigma^2/\mu^3$. In general the k-th moment of the implied lifetime, $\mathbb{E}(T^k)$, $k = 1, 2, \ldots$, can be evaluated numerically by substituting the right-hand side of (4.1) for $\mathbb{P}(T \leq t)$ in

$$\mathbb{E}(T^k) \;=\; k\int_0^\infty t^{k-1}\left(1 - \mathbb{P}(T \leq t)\right)\mathrm{d}t. \tag{4.2}$$

4.2.2 Non-stationary gamma process

Deterioration may also be modelled by a non-stationary gamma process, a monotone stochastic process (cf. Abdel-Hameed, 1975). The process $X = \{X(t),\ t \geq 0\}$ is a non-stationary gamma process with real-valued shape function v, satisfying $v(0) = 0$, and scale parameter $u > 0$, if

- $X(0) = 0$ with probability 1;
- the increment $X(t+h) - X(t)$ is gamma distributed with shape parameter $v(t+h) - v(t)$ and scale parameter u for all $t, h > 0$;

- X has independent increments.

It follows that $X(t)$ is a gamma distributed random variable having shape parameter $v(t)$ and scale parameter u for every $t > 0$. Its probability density function (PDF) is given by

$$f_{X(t)}(x) = \frac{u^{v(t)} x^{v(t)-1}}{\exp(ux)\Gamma(v(t))},$$

where $\Gamma(a) = \int_0^\infty x^{a-1} \exp(-x) \mathrm{d}x$, $a \geq 0$, is the well-known gamma function. Imposing $\mathbb{E}(X(t)) = \mu t^q$ and $\mathrm{Var}(X(t)) = \sigma^2 t^q$ yields shape function $v(t) = \mu^2 t^q / \sigma^2$ and scale parameter $u = \mu/\sigma^2$. Notice that both the shape function and scale parameter depend on μ and σ.

Let T again be the first time X exceeds ρ, then $T = \inf\{t \geq 0 : \ X(t) > \rho\}$. Since the increments of the gamma process are positive, the implied lifetime distribution $F_{GP}(t) := \mathbb{P}(T \leq t)$ is easily derived as

$$\begin{aligned} F_{GP}(t) &= \mathbb{P}(\inf\{\tau \geq 0 : \ X(\tau) > \rho\} \leq t) \\ &= \mathbb{P}(X(t) > \rho) = \frac{\Gamma(v(t), u\rho)}{\Gamma(v(t))}. \end{aligned} \tag{4.3}$$

Here, $\Gamma(a, x) = \int_{u=x}^\infty u^{a-1} \exp(-u) \mathrm{d}u$ is the incomplete gamma function. As for Brownian motion with a non-linear time transformation, the (integer) moments of T can be computed by substituting the right-hand side of (4.3) for $\mathbb{P}(T \leq t)$ in relation (4.2).

4.3 A comparison of Brownian motion and gamma process

One of the main differences between the gamma process (GP) and Brownian motion (BM) is that the former is a monotone stochastic process, whereas the latter is not. Although a monotone process seems to be most appropriate to describe time-dependent deterioration, many scholars have proposed Brownian motion type deterioration models (cf. Doksum and Hóyland, 1992; Whitmore and Schenkelberg, 1997). This may be due to the fact that Brownian motion has continuous sample paths and is related to the normal distribution and as such often facilitates 'easy' computations. The gamma process is a jump process with infinitely many jumps in any finite interval (cf. Van Noortwijk, 2007) and, obviously, it is related to the gamma distribution, which often makes computations somewhat more complex. For instance, although the monotonicity of the gamma process implies that the probability distribution of deterioration and implied lifetime are directly related to each other by relation (4.3), the PDF of the lifetime implied by the gamma process can only be evaluated by numerical differentiation of the right-hand side of (4.3). On the other hand, for Brownian motion an easy to compute analytical expression for the PDF of the implied lifetime is found by differentiating the right-hand side of (4.1).

4.3.1 Case equal parameter values

Let us first consider (non-stationary) Brownian motion and gamma process having equal parameter values. When $\mu \gg \sigma$, the CDF of the lifetime implied by the gamma process, see (4.3), may be approximated by

$$F_{GP}(t) \approx \Phi\left(\frac{\mu t^q - \rho}{\sigma\sqrt{t^q}}\right). \tag{4.4}$$

Park and Padgett (2005) introduce this approximation for a stationary gamma process; the approximation is justified by the central limit theorem.

Also, when $\mu \gg \sigma$ Brownian motion does not have big negative increments and the probability that it does not exceed ρ at time t and does exceed ρ before t, is very small. This means that the second term of equation (4.1) is small. Thus, we can approximate $F_{BM}(t)$ by

$$F_{BM}(t) \approx \Phi\left(\frac{\mu t^q - \rho}{\sigma\sqrt{t^q}}\right). \tag{4.5}$$

By combining (4.4) and (4.5) we have that the implied lifetime distributions can be approximated as follows

$$\begin{aligned} F_{GP}(t) &\approx \Phi\left(\frac{\mu t^q - \rho}{\sigma\sqrt{t^q}}\right) \\ &\approx F_{BM}(t). \end{aligned} \tag{4.6}$$

Note that when t is reasonable large the PDF and CDF of the processes are also approximately equal, since the survival function of the gamma process is directly related to the implied lifetime distribution. However, when t is small, Brownian motion can also take negative values! In order to investigate the quality of these approximations in more detail, we distinguish between two cases: (i) $q \leq 1$, (ii) $q > 1$. Case (i) represents deterioration processes with mean proportional to a concave power function and in case (ii) mean is proportional to a convex power function.

Case (i), $q \leq 1$. From our calculations it appears that the approximations in (4.6) perform very well when $\mu > 10\sigma$. Consequently, for these parameter values the hitting times of the two processes have approximately the same distribution. For instance, Figure 4.1 displays the PDF of the first hitting time of level $\rho = 10$ by Brownian motion and the gamma process. The deterioration parameters are set to $\mu = 5$, $\sigma = 0.5$ and $q = 0.5$. The two densities almost coincide.

Case (ii), $q > 1$. Obviously, the higher the value of q, the better approximation formula (4.6) performs (Brownian motion does not have big negative increments and for the gamma process the approximation of the hitting time distribution is based on the central limit theorem). As a result, both approximations perform well when μ and σ are

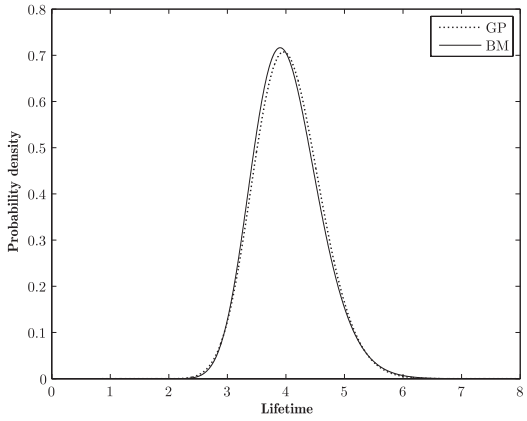

Figure 4.1: Probability density function of the implied lifetime when the failure level is $\rho = 10$ and the deterioration parameters are $\mu = 5$, $\sigma = 0.5$ and $q = 0.5$.

of the same order of magnitude. Figure 4.2 shows the densities of the implied lifetime with parameters $q = 4$, $\mu = \sigma = 1$ and $\rho = 10$. Again, the densities almost coincide. Above

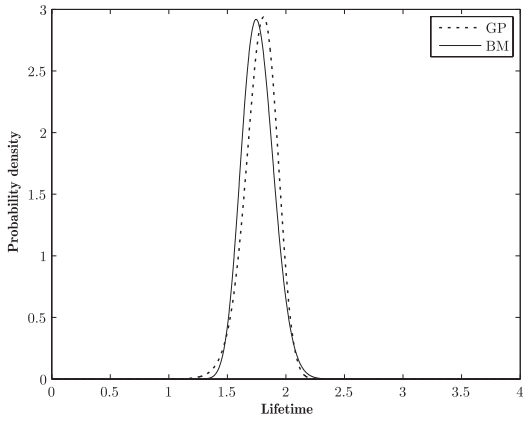

Figure 4.2: Probability density function of the implied lifetime when the failure level is $\rho = 10$ and the deterioration parameters are $\mu = 1$, $\sigma = 1$ and $q = 4$.

reasoning shows that for a specific range of parameter values the implied lifetimes follow the same distribution. It must be said that this range of parameter values corresponds to stochastic processes with little variability. This can be seen by computing the coefficient of variation (CV) of the deterioration process. The CV of X at time $t > 0$ is given by

$$CV(X(t)) = \frac{\sigma}{\mu t^{q/2}}.$$

It is clear that this expression takes small values under the above-mentioned conditions, even when t is small.

A concise comparison of the two processes should also regard cases with high variability. To this end we have to estimate (read: set) the parameters in such a way that we can vary the amount of variability. Parameters are usually obtained by fitting the deterioration process under consideration to sampled data. In general the estimated values are not equal for both processes. Let us assume that the structure subject to deterioration has been inspected at particular points in time. This results in a sample of inspection data or measurements. The parameters of the processes can then be estimated by applying the method of maximum likelihood (see e.g. Çinlar et al., 1997). If we assume that the underlying generating process is either a Brownian motion or a gamma process, and the true parameters of this process satisfy above-mentioned conditions, then estimating the parameters of these processes from inspection data will approximately yield the same values. This is for example the case in Nicolai *et al.* (2007a). However, if the true parameters do not satisfy these conditions, then the characteristics of the estimated processes may be very different. One way to verify this would be by generating data from a Brownian motion or gamma process, and then fit both models to these data. In this study however, we proceed in a different way and adopt the ideas of Pandey *et al.* (2007).

4.3.2 Calibration scheme of Pandey, Yuan and Van Noortwijk

In order to make a fair comparison between random variable and stationary gamma process models, Pandey *et al.* (2007) propose to fit the model parameters in the following way. Firstly, the mean and the coefficient of variation of the implied lifetime, i.e. $\mathbb{E}(T)$ and $CV(T)$, are estimated from a sample of failure times. Next, the (two) model parameters are estimated by means of a two-moment fit. We also consider non-stationary processes and we have to estimate three parameters, viz. μ, σ and q. Van Noortwijk (2007) argues that in practice the value of the power q is often set manually and he gives examples of structures subject to deterioration, for which the value of q ranges between $\frac{1}{8}$ and 4. For given q we can still apply the two-moment fit procedure and so we proceed in the following way:

Scheme 1 (Design of experiments)

(0) Set $\mathbb{E}(T) = 50$ *and* $\rho = 100$.

(1) Vary q *from* 0.5 *to* 4.0 *by* 0.5.

(2) Vary $CV(T)$ *from* 0.1 *to* 0.9 *by* 0.1.

(3) For both processes, estimate μ *and* σ *from* $\mathbb{E}(T)$ *and* $CV(T)$ *by means of a two-moment fit (use relation* (4.2) *with* $k = 1, 2$*).*

That is, we estimate the parameters of Brownian motion and gamma process for different values of q and $CV(T)$. Notice that higher values of $CV(T)$ correspond with higher values of $CV(X(t))$; if the process exhibits little variation, then the lifetime cannot vary much either (and vice versa). Typically, the estimated parameters μ and σ will not be equal for both processes under Scheme 1. For example, when $q = 1$, the mean of the first hitting time of level ρ by a stationary gamma process with rate μ is not equal to ρ/μ, which is the expected lifetime implied by Brownian motion.

In Table 4.1 the parameter estimates of both processes are shown for $q = 1$ and different values of $CV(T)$. The variance parameter σ increases with $CV(T)$ for both processes. However, whereas the drift parameter μ of the gamma process also increases with $CV(T)$, it equals 2 in all cases for Brownian motion. These effects are explained as follows. Obviously, in order to realize highly variable processes σ should be large in both models. Since the distribution of Brownian motion deterioration is symmetric and negative values are possible, the mean rate of the process should just be high enough to compensate slow deterioration and high lifetimes. On the other hand, it cannot be too high since then there will be no big negative increments and then the hitting time is lower than requested. Recall that the parameters μ and σ for the gamma process are obtained by a reparametrization of the shape function and scale parameter and that both parameters affect the scale and shape of the distribution of deterioration. Actually, the shape function is given by $\mu^2 t/\sigma^2$ and the gamma distribution has no left tail (read: is fat on the left-hand side) when its shape parameter is small. That is, when the mean rate of the gamma process is not high enough, the distribution of deterioration is skewed too much to the left and the desired expected lifetime cannot be realized (it will be too high). It also follows from Table 4.1 that the expected deterioration at the mean lifetime is $\mathbb{E}(X(\mathbb{E}(T))) = 459$ when $CV(T) = 0.8$, being much higher than the failure level $\rho = 100$. For Brownian motion it equals $\rho = 100$ for any value of $CV(T)$. Moreover, $CV(X(\mathbb{E}(T))) = CV(T)$, indicating that the variability in deterioration at the mean lifetime is the same as the variability in the lifetime. For the gamma process the variability in deterioration at the mean lifetime appears to be higher than the CV of the lifetime.

From Figure 4.3 it follows that the CDF of Brownian motion at $t = 50$ is above the CDF of the gamma process. That is, Brownian motion takes lower values than the gamma process at $t = 50$. In Figure 4.4 the PDF's of deterioration $X(t)$ are shown for different values of t. As observed by Pandey *et al.* (2007) the shape of the PDF of the gamma process changes over time, as the shape function depends on time. With respect to Brownian motion we observe that it takes negative values, even when $t = 50$.

Table 4.2 shows the parameter values for $q = 4$ and some values of $CV(T) \leq 0.4$. For values of $CV(T)$ higher than 0.4 parameters μ and σ become very large for the gamma process (this may yield some numerical problems) and rather small for Brownian motion. The variance parameter is always much larger than the drift parameter. It is striking that

$CV(T)$		0.2	0.4	0.6	0.8
BM	μ	2	2	2	2
	σ	2.83	5.66	8.49	11.31
GP	μ	2.04	2.20	2.64	4.59
	σ	2.92	6.56	13.05	37.40

Table 4.1: Parameter estimates of Brownian motion and gamma process under Scheme 1 for different values of $CV(T)$ with $q = 1$.

the expected value of Brownian motion (the gamma process) at the expected lifetime is much lower (higher) than the failure level.

$CV(T)$		0.2	0.3	0.4
BM	μ	1.3E−5	9.5E−6	6.5E−6
	σ	3.2E−2	4.9E−2	6.3E−2
GP	μ	1.9E−5	1.4E+4	7.8E+59
	σ	4.3E−2	2.1E+8	3.2E+64

Table 4.2: Parameter estimates of Brownian motion and gamma process under Scheme 1 for different values of $CV(T)$ with $q = 4$.

Figure 4.5 shows the PDF of the lifetimes implied when $CV(T) = 0.4$. Due to the high deterioration rate of the gamma process the implied lifetime can also take low values. On the other hand, the PDF of the lifetime implied by Brownian motion has a heavy right tail and thus the lifetime can take large values. Finally, Table 4.3 shows the parameter

$CV(T)$		0.2	0.4	0.6	0.8
BM	μ	14.21	14.41	14.70	15.06
	σ	3.74	7.39	10.88	14.20
GP	μ	14.29	14.74	15.56	16.91
	σ	3.81	7.93	12.76	18.91

Table 4.3: Parameter estimates of Brownian motion and gamma process under Scheme 1 for different values of $CV(T)$ with $q = \frac{1}{2}$.

values for $q = 0.5$ and different values of $CV(T)$. For both processes the parameters

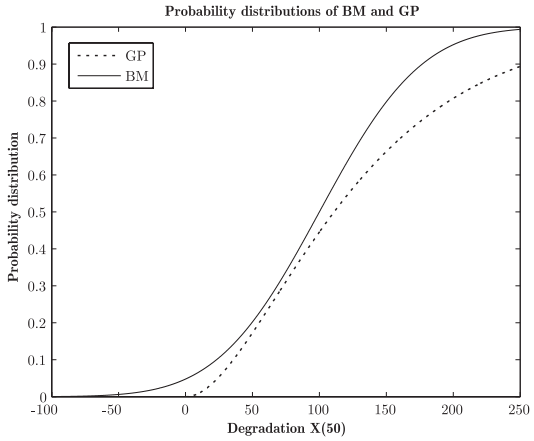

Figure 4.3: Probability distribution of $X(50)$ with $q = 1$ and $CV(T) = 0.6$.

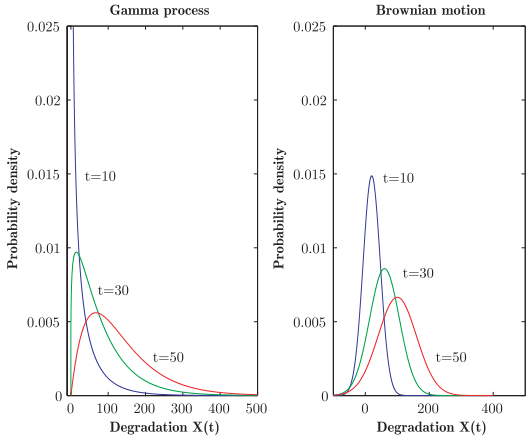

Figure 4.4: Probability density functions of $X(10)$, $X(30)$ and $X(50)$, with $q = 1$ and $CV(T) = 0.6$.

μ and σ now become large when $CV(T)$ becomes large. The expected deterioration is now concave and so deterioration initially grows relative fast. To counterbalance high variability, Brownian motion in this case also has a relative high drift parameter.

4.4 Replacement model

We shall now compare the impact of the two deterioration models on a simple age-replacement policy.

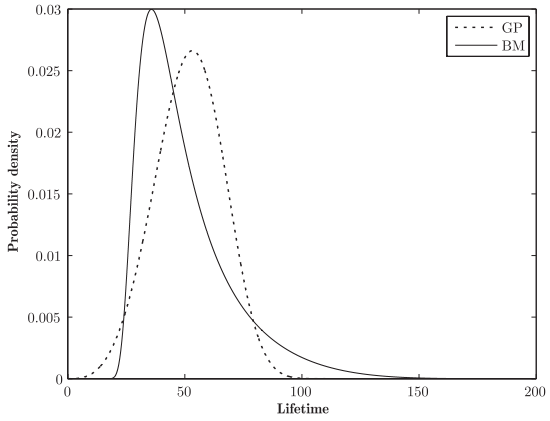

Figure 4.5: Probability density function of the implied lifetime when the failure level is $\rho = 100$, $q = 4$ and $CV(T) = 0.4$.

4.4.1 Assumptions

The main assumptions in our maintenance model are:

1. Single structure subject to random deterioration. Condition is measured on $[0, \infty)$, where 0 denotes the as-good-as-new state.

2. Deterioration is modelled as a, possibly non-stationary, stochastic process $X = \{X(t),\ t \geq 0\}$ starting in zero. We restrict ourselves to two stochastic processes: the gamma process (GP) and Brownian motion (BM). Mean and variance of the deterioration process are proportional to a power function, i.e. $\mathbb{E}(X(t)) = \mu t^q$ and $\mathrm{Var}(X(t)) = \sigma^2 t^q$ with μ, σ and q nonnegative model parameters.

3. Failure is defined as the event in which the deterioration exceeds a fixed failure level ρ.

4. Failures are *not* detected immediately, but only upon intervention, viz. a maintenance action reveals whether deterioration has exceeded the failure level.

5. Replacement restores the structure to as new condition. No other maintenance actions are considered.

6. Replacement cost is a nonnegative non-decreasing function C of deterioration and, in case of failure, also of operating time.

7. The cost criterion is the expected average cost per unit time.

4.4.2 Cost function

By assuming that failures are not detected immediately, the cost of a failure replacement becomes very important in our model. To see why, assume that the cost of a failure replacement is given by a constant c_F. Since the expected average cost per unit time tends to zero as the replacement age tends to infinity, it can be minimized by doing *no* maintenance at all. This is undesirable. In general an engineering structure fulfills a valuable task and a failure can bring about time-dependent or condition-dependent costs. Thus, in principle a 'no-maintenance' or 'laissez-faire' strategy should not be optimal.[2]

Let us now define a cost function C which penalizes failures in a different way. The cost of preventive replacement is a function of measured deterioration. Next, the cost of failure replacement is a function of measured deterioration and of operating time. Note that a replacement is called 'failure replacement' if the measured deterioration at maintenance exceeds the failure level. That is, if maintenance is done at time t_0, then the replacement is a failure replacement if $X(t_0) > \rho$ and otherwise it is considered a preventive one (even if ρ was exceeded before, which is possible if we model deterioration by a Brownian motion type process). Denote operating time by t and measured deterioration at intervention by x. Now the exact specification of the cost function $C(t, x)$ is

$$C(t,x) = \begin{cases} c_P, & x \leq 0, t \geq 0 \\ c_P + c_1 x, & 0 < x \leq \rho, t \geq 0 \\ c_F + c_2 x + c_3 t, & x > \rho, t \geq 0 \end{cases}, \tag{4.7}$$

where $c_P, c_F, c_1, c_2, c_3 > 0$ and $c_P + c_1\rho \leq c_F + c_2\rho$. Note that the cost of preventive replacement is c_P if measured deterioration is negative. This can only occur if deterioration is modelled as a Brownian motion.

4.4.3 Age replacement

The standard age replacement policy dictates a fixed replacement interval of length t_0 and corrective replacement is done immediately upon failure. However, in our model failures are only detected upon intervention and thus age replacement implies that the replacement cycle always has length t_0. The expected maintenance cost in this cycle is

[2]In the standard age-replacement model, where a structure is immediately replaced upon failure, the strategy 'plan no preventive replacement' is cost-effective if the structure's lifetime is exponentially distributed. This implies that the system is only replaced upon failure. Pandey *et al.* (2007) report that this strategy can also be optimal if the structure's lifetime is highly uncertain.

given by

$$\begin{aligned}\mathbb{E}(C(t_0, X(t_0))) &= \int_0^\infty C(t_0, x) f_{X(t_0)}(x)\mathrm{d}x \\ &= c_P \mathbb{P}(X(t_0) \leq \rho) + (c_F + c_3 t_0)\mathbb{P}(X(t_0) > \rho) \\ &\quad + c_1 \int_0^\rho x f_{X(t_0)}(x)\mathrm{d}x + c_2 \int_\rho^\infty x f_{X(t_0)}(x)\mathrm{d}x \\ &= c_P + c_1\rho + (c_F + c_2\rho + c_3 t_0 - c_P - c_1\rho)\mathbb{P}(X(t_0) > \rho) \\ &\quad - c_1 \int_0^\rho \mathbb{P}(X(t_0) \leq x)\mathrm{d}x + c_2 \int_\rho^\infty \mathbb{P}(X(t_0) > x)\mathrm{d}x. \end{aligned} \tag{4.8}$$

Note that these equations are valid when X is either a Brownian motion or a gamma process. Observe that $\mathbb{E}(C(0, X(0))) = c_P$ and $\mathbb{E}(C(t_0, X(t_0))) \to c_F + c_3 t_0 + c_2 \mathbb{E}(X(t_0))$ as $t_0 \to \infty$, implying that the expected average cost per unit time (also referred to as the mean cost rate) tends to $c_3 + c_2 \mu t_0^{q-1}$ as $t_0 \to \infty$. The optimal age replacement can now be found by minimizing $\mathbb{E}(C(t_0, X(t_0)))/t_0$ with respect to t_0.

4.4.4 Numerical experiments

Numerical experiments with the age replacement policy have been done to assess the effect of Brownian motion and gamma process on the maintenance decisions. We have set $c_1 = c_2 = 0$, $c_3 = 5$, $c_P = 10$ and $c_F = 50$. That is, the cost of preventive replacement is fixed and the cost of corrective replacement depends on the operating time of the structure. The deterioration parameters are set according to Scheme 1.

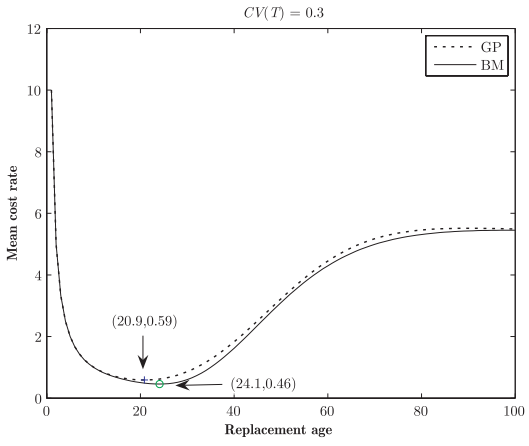

Figure 4.6: Expected average cost per unit time versus the replacement age for gamma process and Brownian motion with $q = 1$ and $CV(T) = 0.3$.

In Figure 4.6 the mean cost rate is plotted against the replacement age for $q = 1$ and $CV(T) = 0.3$. The cost rates do not differ much, but the optima are somewhat different. The optimal mean cost rate for Brownian motion is somewhat lower (0.46 vs 0.59).

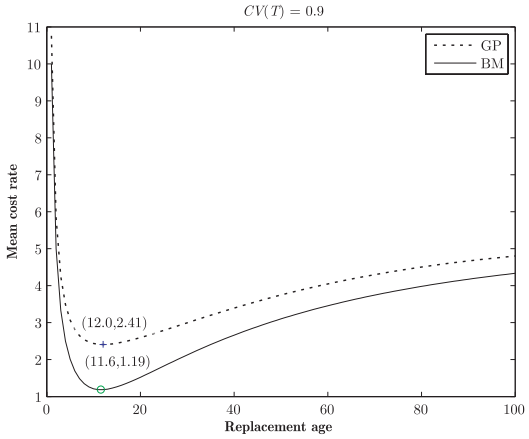

Figure 4.7: Expected average cost per unit time versus the replacement age for gamma process and Brownian motion with $q = 1$ and $CV(T) = 0.9$.

Figure 4.7 shows the same functions for $CV(T) = 0.9$. The mean cost rate of Brownian motion is well below that of the gamma process. Apparently, the probability that $X(t)$ exceeds ρ is higher for the latter process, for any value of t. This may be explained by the high drift parameter of the gamma process. The optimal replacement age is almost equal for both processes. We do not observe that the optimal replacement age tends to infinity when the variability in deterioration is high, which was found in Pandey *et al.* (2007).

In Figures 4.8 and 4.9 the optimal replacement age (left) and mean cost rate (right) are plotted versus $CV(T)$ with $q = 1$ and $q = 2$. The mean cost rate is an increasing function of $CV(T)$. It is in all cases higher in the gamma process model. The optimal replacement ages both decrease with $CV(T)$ and are almost the same when the variability is very low or very high ($q = 1$), but in between it is higher for Brownian motion.

4.5 Conclusions

We have compared two stochastic processes that are often used to model deterioration of engineering structures, viz. Brownian motion with a non-linear time transformation and the non-stationary gamma process. We find that both processes do not differ much when the uncertainty associated with deterioration is low or the drift of the process is very large. As a consequence their hitting times, which can be used to represent the

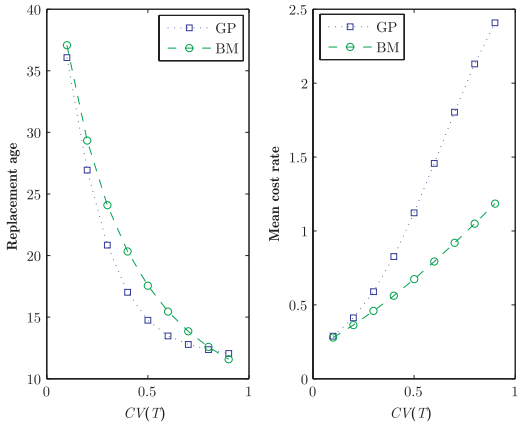

Figure 4.8: Optimal replacement age versus $CV(T)$ (left) and optimal expected average cost per unit time (right) for gamma process and Brownian motion with $q = 1$.

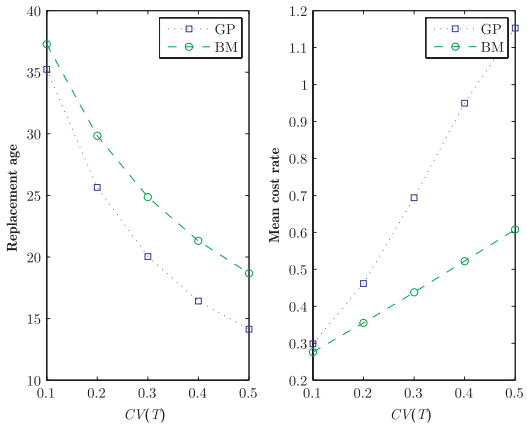

Figure 4.9: Optimal replacement age versus $CV(T)$ (left) and optimal expected average cost per unit time (right) for Brownian motion and the gamma process with $q = 2$.

lifetime of a structure, follow approximately the same distribution. When the variability in the lifetime is high, the parameter values of the gamma process become very large and this may give some numerical difficulties. On the other hand, Brownian motion takes large negative values and for this reason it should not be used when the variability in deterioration is high.

We have also studied a simple age replacement policy for structures subject to time-dependent deterioration, where failures are not observed immediately. The two deterioration models lead to different maintenance decisions. Since in the gamma process model

early failures are more likely, the replacement age is lower than it is for the Brownian motion model when variability in deterioration is fairly high. Moreover, the expected average cost per unit time is higher for the gamma process. The gamma process appears to be the safer option when modelling deterioration.

Chapter 5

Approximating the randomized hitting time of a non-stationary gamma process*

Abstract

The non-stationary gamma process is a non-decreasing stochastic process with independent increments. By this monotonic behavior this stochastic process serves as a natural candidate for modelling time-dependent phenomena such as degradation. In the field of structural reliability and condition-based maintenance the first time such a process exceeds a random threshold is used as a model for the lifetime of a device or for the random time between two successive imperfect maintenance actions. Therefore there is a need to investigate in detail the cumulative distribution function (CDF) of this so-called randomized hitting time. We first relate the CDF of the (randomized) hitting time of a non-stationary gamma process to the CDF of a related hitting time of a stationary gamma process. Even for a stationary gamma process this CDF has in general no elementary formula and its evaluation is time-consuming. Hence two approximations are proposed in this chapter and both have a clear probabilistic interpretation. Numerical experiments show that these approximations are easy to evaluate and their accuracy depends on the scale parameter of the non-stationary gamma process. Finally, we also consider some special cases of randomized hitting times for which it is possible to give an elementary formula for its CDF.

*This chapter is based on Frenk and Nicolai (2007)

5.1 Introduction

The non-stationary gamma process is a monotone continuous-time non-homogeneous Markov process with independent increments. The increments are gamma distributed with time-dependent shape function and an identical scale parameter. It is a jump process and the number of jumps in any time interval is infinite with probability one (Lawless and Crowder, 2004). These properties make the gamma process a suitable candidate to model the temporal variability in monotone phenomena. Van Noortwijk (2007) gives an overview of the application of gamma processes within maintenance. Within this field these processes are used to describe time-dependent deterioration such as wear, creep and corrosion. Another application of the gamma process outside the field of maintenance, is the aggregate claim process within insurance mathematics Dufresne *et al.* (1991).

Van Noortwijk indicates in his review that many authors have studied the first time a gamma process exceeds a fixed threshold value. However, only a few consider the first time a random threshold value is exceeded. In the remainder of this chapter such a time is called a randomized hitting time. The first to propose a randomized hitting time within structural reliability and maintenance was Abdel-Hameed (1975). He uses the gamma process to model deterioration, in particular wear, and the associated randomized hitting time serves as a model for the lifetime of a device. This means that the cumulative distribution function (CDF) of the lifetime of the device is the CDF of a randomized hitting time associated with a given gamma process. However, as indicated by Van Noortwijk, authors proposing this randomized hitting time model do mostly not perform numerical experiments. This is probably due to the complicated structure of the CDF being a two-dimensional integral. Since computing the CDF of this randomized hitting time is a subprocedure, which for some maintenance optimization models needs to be repeated under different parameters, the numerical optimization of such a model is mostly time-consuming. Hence there is a need for a fast numerical procedure giving reliable numerical outcomes for this CDF. As an example we mention a model using randomized hitting times introduced by Nicolai and Frenk (2006), which will be the topic of Chapter 6. The duration between two maintenance actions is modelled by the time at which the gamma process (representing the deterioration process of a steel structure) exceeds a random reduction resulting from the last imperfect maintenance action. The problem of selecting maintenance actions resulting in the lowest expected cost over a finite horizon is then formulated as a stochastic dynamic programming model. For this model the optimal policy needs to be computed. Due to the above considerations the main purpose of this chapter is to investigate in detail the CDF of a randomized hitting time of a (non-)stationary gamma process.

First of all, we show in Section 5.2 that without loss of generality we may restrict ourselves to the CDF of a randomized hitting time associated with a so-called standard

gamma process. In this section we also study in detail the CDF of the fractional and integer part of a randomized hitting time and extend one of the results discussed in Steutel and Thiemann (1990). At the same time we simplify the proof technique for this result. Applying now the theoretical results of this section we propose in Section 5.3 two approximations for the CDF of a randomized hitting time. The first one is based on the approximative assumption that the fractional and the integer part of the randomized hitting time are independent and that the CDF of the fractional part is uniformly distributed on $(0, 1)$. To justify these approximative assumptions (especially the latter one) we apply the results derived in Section 5.2. The second approximation is based on the assumption that the sample path of a stationary gamma process can be replaced by a piecewise linear sample path that coincides with the original sample path at integer points. As in both approximations the true CDF at integer time points is required, we also give a method for computing these probabilities. In Section 5.4 we derive for some random thresholds having a gamma-type CDF elementary formulas for the CDF of the associated randomized hitting times. Finally, in Section 5.5 numerical experiments are done to assess the quality of both approximations proposed in Section 5.3. We also compare the computation time for the construction of the approximations and the true CDF's.

5.2 On randomized hitting times of a non-stationary gamma process

In this section we derive some properties of the hitting time of a non-stationary gamma process exceeding a nonnegative random threshold value. To introduce the definition of a gamma process we first observe (Steutel and van Harn, 2004) that a gamma density with scale parameter $\lambda > 0$ and shape parameter $\beta > 0$ is given by

$$f(x) = \Gamma(\beta)^{-1}\lambda^{\beta}x^{\beta-1}\exp(-\lambda x)1_{(0,\infty)}(x)$$

with

$$\Gamma(\beta) := \int_0^{\infty} x^{\beta-1}\exp(-x)\mathrm{d}x$$

the well-known gamma function. In the rest of this chapter such a CDF is denoted by gamma(β, λ). Also we mean by $X \sim F$ that the random variable X has CDF F.

Definition 5.1 *Let $\lambda > 0$ and $v : [0, \infty) \to [0, \infty)$ an increasing, right-continuous function satisfying $v(0) = 0$. The stochastic process $X_{v,\lambda} = \{X_{v,\lambda}(t) : t \geq 0\}$ is called a gamma process with shape function v and scale parameter $\lambda > 0$ if*

1. *$X_{v,\lambda}(0) = 0$ with probability 1 and $X_{v,\lambda}$ is a cadlag process.*

2. *The stochastic process $X_{v,\lambda}$ has independent increments.*

3. *The random variable $X_{v,\lambda}(s) - X_{v,\lambda}(t)$, $s > t$ has a gamma distribution with shape parameter $v(s) - v(t)$ and scale parameter $\lambda > 0$.*

If the function v is linear a gamma process $X_{v,\lambda}$ is called stationary, otherwise it is called non-stationary. A stationary gamma process with v having slope 1 and scale parameter λ equal to 1 is called standard and denoted by X. In Protter (1992) it is shown that there exists a unique modification of the standard gamma process which is càdlàg (right continuous sample paths having left-hand limits). We will always use this modification.

For any (non-stationary) gamma process we are now interested in the so-called hitting time $T_{v,\lambda}(r)$ of level $r > 0$ given by

$$T_{v,\lambda}(r) := \inf\{t \geq 0 : X_{v,\lambda}(t) > r\}. \tag{5.1}$$

If the gamma process is standard the hitting time in (5.1) is denoted by $T(r)$. In the next chapter we are interested in the randomized hitting time $T_{v,\lambda}(R)$ with R a nonnegative random variable independent of the (non-stationary) gamma process. This random variable represents the random duration between two imperfect maintenance actions. In particular, the (non-stationary) gamma process models the deterioration of the structure under consideration and the random variable R having a non-defective CDF G_R satisfying $G_R(0) = 0$ denotes the random reduction in damage due to an imperfect maintenance action. We assume throughout this chapter that the random variable R is independent of the deterioration process $X_{v,\lambda}$. To relate the properties of the (randomized) hitting time to the hitting time of a standard gamma process we observe by Definition 5.1 that

$$X_{v,\lambda}(t) \stackrel{d}{=} \lambda^{-1} X_{v,1}(t) \stackrel{d}{=} \lambda^{-1} X(v(t)) \tag{5.2}$$

where $\stackrel{d}{=}$ is used to indicate that two random variables have the same CDF. Using relation (5.2) the next result holds.

Lemma 5.2 *If $\lambda > 0$ and the shape function v is strictly increasing and continuous satisfying $v(0) = 0$ and $v(\infty) = \infty$, then*

$$T_{v,\lambda}(R) \stackrel{d}{=} v^{\leftarrow}(T(\lambda R)) \tag{5.3}$$

where $v^{\leftarrow}$ denotes the inverse function of v.

Proof. By relations (5.1) and (5.2) we obtain

$$T_{v,\lambda}(R) \stackrel{d}{=} \inf\{t \geq 0 : X(v(t)) > \lambda R\}. \tag{5.4}$$

Since the function v is strictly increasing and continuous with range $[0, \infty)$ its inverse $v^{\leftarrow}$ is also strictly increasing and continuous and satisfies $v^{\leftarrow}(v(t)) = t$ for every $t \geq 0$. This shows by relation (5.4) that

$$T_{v,\lambda}(R) \stackrel{d}{=} \inf\{v^{\leftarrow}(t) : X(t) > \lambda R\} = v^{\leftarrow}(\inf\{t \geq 0 : X(t) > \lambda R\})$$

and we have verified the result. □

For a gamma process with continuous shape function it is well-known that it is an increasing jump process with a countably infinite number of jumps over any finite interval (see Lawless and Crowder, 2004; Bertoin, 1996) and so the overshoot at any given level is a non-degenerate random variable. If the overshoot $W_{v,\lambda}(r)$ of level r of a non-stationary gamma process is given by

$$W_{v,\lambda}(r) := X_{v,\lambda}(T_{v,\lambda}(r)) - r \tag{5.5}$$

we obtain by Lemma 5.2 and relation (5.2) that

$$W_{v,\lambda}(R) \stackrel{d}{=} \lambda^{-1} X(T(\lambda R)) - R = \lambda^{-1} W(\lambda R) \tag{5.6}$$

with $W(\lambda R)$ the overshoot of a standard gamma process beyond level λR. As for the randomized hitting time this shows that the overshoot of a non-stationary gamma process can also be reduced to the overshoot of a standard gamma process.

To investigate in detail the CDF of the hitting time $T_{v,\lambda}(R)$ it follows by Lemma 5.2 introducing $H_R(t) := \mathbb{P}\{T(R) \leq t\}$ that

$$\mathbb{P}\{T_{v,\lambda}(R) \leq t\} = H_{\lambda R}(v(t)). \tag{5.7}$$

Hence, to compute or approximate the CDF of the random variable $T_{v,\lambda}(R)$ for any non-stationary gamma process, it is therefore sufficient to compute or approximate the CDF H_R. By relation (5.7) it is clear for $A_{\lambda R} : [0, \infty) \to [0, 1]$ an approximation of the CDF of $T(\lambda R)$ that the approximation $A_{v,\lambda}$ of the CDF of the random variable $T_{v,\lambda}(R)$ is given by

$$A_{v,\lambda}(t) := A_{\lambda R}(v(t)). \tag{5.8}$$

For v continuous and increasing satisfying $v(0) = 0$ and $v(\infty) = \infty$ relation (5.8) implies

$$\|A_{v,\lambda} - H_{v,\lambda}\|_\infty = \|A_{\lambda R} - H_{\lambda R}\|_\infty \tag{5.9}$$

with $H_{v,\lambda}$ denoting the CDF of $T_{v,\lambda}(R)$ and $\|h\|_\infty := \sup_{0 \leq t < \infty} |h(t)|$ the well-known supnorm. Since the nonnegative random variable R is by definition independent of the gamma process and its CDF satisfies $G_R(0) = 0$ it follows by conditioning on the random variable R that

$$H_R(t) = \mathbb{P}\{T(R) \leq t\} = \mathbb{P}\{X(t) > R\} = \int_0^\infty \mathbb{P}\{X(t) > r\} dG_R(r) \tag{5.10}$$

for every $t \geq 0$. Moreover, by conditioning on the random variable $X(t)$ with CDF F_t we obtain for G_R continuous the equivalent representation

$$H_R(t) = \int_0^\infty \mathbb{P}\{R < x\} dF_t(x) = \mathbb{E}(G_R(X(t))). \tag{5.11}$$

Note that Equation (5.10) is exactly the one that was proposed in Abdel-Hameed (1975). In addition to a gamma deterioration process, Abdel-Hameed assumed a random failure level as well. Since the gamma process is increasing and continuous in probability relation (5.10) and a standard application of Lebesque's dominated convergence theorem (see e.g. Billingsley, 1979) imply that the CDF H_R is continuous. Using relation (5.11) one can also derive for random variables R having a gamma type CDF an elementary formula for the CDF H_R. This will be done in Section 5.4. For general CDF's this is not possible and so we propose in the next section two elementary approximations. To justify these approximations we first investigate some properties of random variables related to the hitting time.

Let $\lfloor x \rfloor$ be the largest integer not exceeding x for any $x \geq 0$ and denote by $\mathcal{F}(x)$ its fractional part given by $\mathcal{F}(x) := x - \lfloor x \rfloor$. Clearly

$$T(R) = \lfloor T(R) \rfloor + \mathcal{F}(T(R)). \tag{5.12}$$

For the integer part $\lfloor T(R) \rfloor$ it is easy to show the following result.

Lemma 5.3 *If $N := \{N(t) : t \geq 0\}$ is a Poisson process with arrival rate 1, then $\lfloor T(R) \rfloor \stackrel{d}{=} N(R)$.*

Proof. Since for a standard gamma process

$$X(n+1) \stackrel{d}{=} \sum\nolimits_{k=1}^{n+1} Y_k$$

with $Y_k, 1 \leq k \leq n+1$ independent and exponentially distributed with scale parameter 1 we obtain by the continuity of the CDF of the random variable $T(R)$ that

$$\begin{aligned} \mathbb{P}\{\lfloor T(R) \rfloor \leq n\} &= \mathbb{P}\{T(R) < n+1\} = \mathbb{P}\{T(R) \leq n+1\} \\ &= \mathbb{P}\{X(n+1) > R\} = \mathbb{P}\{N(R) \leq n\}. \end{aligned} \tag{5.13}$$

Hence the desired result follows. □

By the independence of the random variable R and the standard gamma process X we obtain by Lemma 5.3 and the well-known properties of a Poisson process that

$$\mathbb{P}\{\lfloor T(R) \rfloor = n\} = \frac{1}{n!} \mathbb{E}\left(R^n \exp(-R)\right). \tag{5.14}$$

for every $n \in \mathbb{N} \cup \{0\}$. In the next lemma we give an expression for the CDF of $\mathcal{F}(T(R))$. For this expression we introduce the functions $q_t : (0, \infty) \to \mathbb{R}$, $0 < t < 1$ given by

$$q_t(r) := 1 - F_t(r) - \int_r^\infty 1 - F_t(x)dx. \tag{5.15}$$

Lemma 5.4 *If R is a nonnegative random variable with CDF G_R satisfying $G_R(0) = 0$ and this random variable is independent of the standard gamma process X, then*

$$\mathbb{P}\{\mathcal{F}(T(R)) \leq t\} = t + \mathbb{E}\left(q_t(R)\right)$$

for every $0 < t < 1$.

Proof. By the definition of the random variable $\mathcal{F}(T(R))$ we obtain

$$\{\mathcal{F}(T(R)) \leq t\} = \cup_{k=0}^\infty \{k \leq T(R) \leq k + t\}. \tag{5.16}$$

Since the CDF of $T(R)$ is continuous and so $\{k \leq T(R) \leq k + t\} \stackrel{a.s}{=} \{k < T(R) \leq k + t\}$ it follows for every $k \in \mathbb{N} \cup \{0\}$ that

$$\{k \leq T(R) \leq k + t\} \stackrel{a.s}{=} \{X(k) \leq R < X(k + t)\}.$$

This yields by relation (5.16)

$$\{\mathcal{F}(T(R)) \leq t\} \stackrel{a.s}{=} \cup_{k=0}^\infty \{X(k) \leq R < X(k + t)\}$$

and conditioning on R implies

$$\mathbb{P}\{\mathcal{F}(T(R)) \leq t\} = \int_0^\infty \mathbb{P}\{\cup_{k=0}^\infty \{X(k) \leq r < X(k + t)\}\} dG_R(r). \tag{5.17}$$

Since $X(k+t) \stackrel{d}{=} X(k)+Y(t)$ with Y a standard gamma process independent of X and the CDF of $X(k)$ is the k-fold convolution of an exponential distribution with scale parameter 1, we obtain for every $r > 0$

$$\begin{array}{rcl} \mathbb{P}\{\cup_{k=0}^\infty \{X(k) \leq r < X(k + t)\}\} & = & \sum_{k=0}^\infty \mathbb{P}\{X(k) \leq r < X(k) + Y(t)\} \\ & = & 1 - F_t(r) + \int_0^r \left(1 - F_t(r - x)\right) dU(x) \end{array} \tag{5.18}$$

with $U(x) = \sum_{k=1}^\infty F^{k*}(x) = x$ the renewal function associated with a Poisson process having arrival rate 1. Using now

$$\int_0^\infty 1 - F_t(x)dx = \mathbb{E}(X(t)) = t$$

and relations (5.17) and (5.18) the desired result follows. □

To rewrite the representation of the function q_t into a more suitable form we introduce a beta CDF on $(0,1)$ with parameters $\alpha, \beta > 0$. Its density is given by

$$b_{\alpha,\beta}(x) := \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} 1_{(0,1)}(x) \tag{5.19}$$

and the CDF itself is denoted by beta(α, β). If $\beta > \alpha > 0$ and $\tau, \lambda > 0$ it is shown on page 261 of Whittaker and Watson (1958) (see also Steutel, 1971) that

$$\left(\frac{\lambda}{\lambda+\tau}\right)^{\alpha} = \frac{\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta-\alpha)} \int_0^1 \left(\frac{\lambda}{\lambda+\tau x}\right)^{\beta} x^{\alpha-1}(1-x)^{\beta-\alpha-1} dx. \tag{5.20}$$

and this result leads to a well-known probabilistic interpretation of a random variable having a gamma CDF with non-integer shape parameter. Although this result seems to be well-known for the special case $\beta = 1$) (Devroye, 1986; Steutel and van Harn, 2004; Robert and Casella, 1999), we could not find a detailed proof of this. Hence in the next lemma we list a short proof based on relation (5.20).

Lemma 5.5 *If $0 < t < 1$ and Y and V are independent nonnegative random variables with $Y \sim$ gamma$(\beta, \lambda), \beta > 0$ and $V \sim$ beta$(t\beta, (1-t)\beta)$, then $VY \sim$ gamma$(t\beta, \lambda)$.*

Proof. For every $\tau > 0$ it follows by the independence of V and Y that

$$\begin{aligned}
\mathbb{E}\left(\exp(-\tau VY)\right) &= \frac{\Gamma(\beta)}{\Gamma(t\beta)\Gamma((1-t)\beta)} \int_0^1 \mathbb{E}\left(\exp(-\tau v Y)\right) v^{t\beta-1}(1-v)^{(1-t)\beta-1} dv \\
&= \frac{\Gamma(\beta)}{\Gamma(t\beta)\Gamma(\beta-t\beta)} \int_0^1 \left(\frac{\lambda}{\lambda+\tau v}\right)^{\beta} v^{t\beta-1}(1-v)^{\beta-t\beta-1} dv
\end{aligned}$$

Applying relation (5.20) we obtain using $\alpha = t\beta < \beta$ that $\mathbb{E}(\exp(-\tau VY)) = (\frac{\lambda}{\lambda+\tau})^{t\beta}$ and this is the probability Laplace-Stieltjes transform (pLSt) of a gamma distribution with scale parameter λ and shape parameter $t\beta$. □

To rewrite $\mathbb{E}\left(q_t(R)\right)$ in a more suitable form we also introduce a Pareto distribution on $(0,\infty)$ with parameter $\beta > 0$. Observe a Pareto(β) CDF on $(0,\infty)$ with parameter $\beta > 1$ has density (Steutel and van Harn (2004))

$$f(x) = (\beta-1)(1+x)^{-\beta} 1_{(0,\infty)}(x) \tag{5.21}$$

and this CDF is denoted by par(β).

Lemma 5.6 *For every $0 < t < 1$ it follows for any nonnegative random variable R and $Z \sim$ par(2) independent of R that*

$$\mathbb{E}\left(q_t(R)\right) = \pi^{-1}\sin(\pi t)\mathbb{E}\left(\exp(-R(Z+1))Z^{1-t}\right).$$

Proof. It is sufficient to show that

$$q_t(r) = \pi^{-1} \sin(\pi t) \int_0^\infty \exp(-r(z+1)) z^{1-t} (1+z)^{-2} dz. \tag{5.22}$$

for every $0 < t < 1$. Since $X(t) \sim \text{gamma}(t, 1)$ it follows by relation (5.15) and Lemma 5.5 (take $\beta = 1$ and $\lambda = 1$) that

$$q_t(r) = \mathbb{P}\{VY > r\} - \int_r^\infty \mathbb{P}\{VY > x\} dx \tag{5.23}$$

with $Y \sim \text{gamma}(1, 1)$ and $V \sim \text{beta}(t, 1-t)$ independent random variables. By Tonelli's theorem (Kawata, 1972)

$$\begin{aligned} \int_r^\infty \mathbb{P}\{VY > x\} dx &= \int_r^\infty \int_0^1 \mathbb{P}\{vY > x\} b_{t,1-t}(v) dv dx \\ &= \int_0^1 \int_r^\infty \mathbb{P}\{vY > x\} dx b_{t,1-t}(v) dv \end{aligned}$$

and this implies by relation (5.23) that

$$q_t(r) = \int_0^1 \left(\mathbb{P}\{vY > r\} - \int_r^\infty \mathbb{P}\{vY > x\} dx \right) b_{t,1-t}(v) dv. \tag{5.24}$$

Since $Y \sim \text{gamma}(1, 1)$ it follows

$$\mathbb{P}\{vY > r\} - \int_r^\infty \mathbb{P}\{vY > x\} dx = (1-v) \exp(-rv^{-1})$$

for every $0 < v < 1$ and so by relations (5.19) and (5.24) we obtain

$$\begin{aligned} q_t(r) &= \tfrac{\Gamma(1)}{\Gamma(t)\Gamma(1-t)} \int_0^1 v^{t-1} (1-v)^{1-t} \exp(-rv^{-1}) dv \\ &= \tfrac{1}{\Gamma(t)\Gamma(1-t)} \int_1^\infty \exp(-rz)(z-1)^{1-t} z^{-2} dz \\ &= \tfrac{1}{\Gamma(t)\Gamma(1-t)} \int_0^\infty \exp(-r(z+1)) z^{1-t} (1+z)^{-2} dz. \end{aligned} \tag{5.25}$$

Using now Euler's reflection formula for the gamma function (see page 256 of Abramowitz and Stegun, 1972) or Whittaker and Watson, 1958) given by $\Gamma(t)\Gamma(1-t) = \pi(\sin \pi t)^{-1}$ for $0 < t < 1$ the desired representation for $q_t(r)$ listed in relation (5.22) is shown. □

Combining Lemma 5.4 and 5.6 we immediately obtain

$$\mathbb{P}\{\mathcal{F}(T(R)) \le t\} = t + \pi^{-1} \sin(\pi t) \mathbb{E}(\exp(-R(Z+1)) Z^{1-t}). \tag{5.26}$$

We will now show that the last nonnegative term in relation (5.26) is small for $\mathbb{E}(\exp(-\mathbb{R}))$ small. Hence we may conclude in this case that the random variable $\mathcal{F}(T(R))$

is approximately uniformly distributed. We will use this observation in the next section to justify an approximation for the CDF of $T(R)$. Since Z and R are random variables on $(0,\infty)$ it follows that

$$\exp(-R(Z+1))Z^{1-t} \leq \exp(-R)Z^{1-t}.$$

By the independence of Z and R and $Z \sim par(2)$ this implies for every $0 < t < 1$ that

$$\mathbb{E}(\exp(-R(Z+1))Z^{1-t}) \leq \mathbb{E}(\exp(-R))\mathbb{E}(Z^{1-t}) = \mathbb{E}(\exp(-R))O(t^{-1}).$$

Hence by (5.26) we obtain

$$0 \leq \mathbb{P}\{\mathcal{F}(T(R)) \leq t\} - t = O\left(\frac{\sin(\pi t)}{\pi t}\right)\mathbb{E}(\exp(-R)). \tag{5.27}$$

A related upperbound for R a degenerate random variable is derived in Steutel and Thiemann (1990) using a completely different technique. This upperbound can be derived considering relation (5.22). In Steutel and Thiemann (1990) the inversion formula for Laplace transforms is used together with an appropriate choice of the closed contour in the associated complex contour integral. To compute the integer moments of $\mathcal{F}(T(R))$ we need to evaluate for every $k \in \mathbb{N} \cup \{0\}$ the function $M_k : \mathbb{R} \to \mathbb{R}$ given by

$$M_k(a) := \int_0^1 \exp(at)t^k \sin(\pi t)dt. \tag{5.28}$$

In the next lemma the function M_{-1} denotes the zero function.

Lemma 5.7 *It follows* $M_0(a) = \frac{\pi}{a^2+\pi^2}(\exp(a)+1)$ *and*

$$M_k(a) = \frac{\pi \exp(a)}{a^2+\pi^2} - \frac{2ka}{a^2+\pi^2}M_{k-1}(a) - \frac{k(k-1)}{a^2+\pi^2}M_{k-2}(a) \tag{5.29}$$

for every $k \in \mathbb{N}$.

Proof. It is easy to check by differentiation that the antiderivative of the function $t \mapsto \exp(at)\sin(\pi t)$ on $(0,\infty)$ is given by the function

$$t \mapsto \frac{\exp(at)}{a^2+\pi^2}(a\sin(\pi t) - \pi\cos(\pi t)).$$

and so the expression for $M_0(a)$ is verified. To check the recurrence relation (5.29) we observe by the first part that

$$\pi(\exp(a)+1) = (a^2+\pi^2)M_0(a). \tag{5.30}$$

Since for every $k \in \mathbb{N}$ it follows that $M_k(a) = M_0^{(k)}(a)$ with $M_0^{(k)}$ denoting the kth derivative of the function M_0 we obtain by differentiation of the identity in (5.30) the desired result. □

Using a package such as Maple and the above recurrence relation it is possible to give an analytical expression for $M_k(a)$. In the next lemma we give an expression for the first moment of $\mathcal{F}(T(R))$.

Lemma 5.8 *If R is a nonnegative random variable with CDF G_R satisfying $G_R(0) = 0$ and this random variable is independent of the standard gamma process X, then*

$$\mathbb{E}(\mathcal{F}(T(R)) = \frac{1}{2} - \mathbb{E}\left(\frac{\exp(-R(Z+1))(Z+1)}{(\ln(Z))^2 + \pi^2}\right)$$

with Z independent of R and $Z \sim par(2)$.

Proof. Since for any random variable Y on $(0,1)$ it is well-known that $\mathbb{E}(Y) = \int_0^1 \mathbb{P}\{Y > t\}dt$ we obtain by relation (5.26)

$$\mathbb{E}(\mathcal{F}(T(R)) = \frac{1}{2} - \pi^{-1}\int_0^1 \sin(\pi t)\mathbb{E}(\exp(-R(Z+1)Z^{1-t})dt. \tag{5.31}$$

Applying Tonelli's theorem and relation (5.28) yields

$$\int_0^1 \sin(\pi t)\mathbb{E}(\exp(-R(Z+1)Z^{1-t})dt = \mathbb{E}\left(\exp(-R(Z+1))ZM_0(-\ln(Z))\right) \tag{5.32}$$

and by Lemma 5.7 the desired result follows. □

By relation (5.12) and Lemma 5.3 we immediately obtain

$$\mathbb{E}(T(R)) = \mathbb{E}(R) + \frac{1}{2} - \mathbb{E}\left(\frac{\exp(-R(Z+1))(Z+1)}{(\ln(Z))^2 + \pi^2}\right). \tag{5.33}$$

To derive a simple upper and lower bound on the expectation of $T(R)$ or $\mathcal{F}(T(R))$ we observe for any nonnegative random variable Z that $(\ln(Z))^2 + \pi^2 \geq \pi^2$. This shows by Lemma 5.8 that

$$\frac{1}{2} - \pi^{-2}\mathbb{E}\left(\exp(-R(Z+1))(Z+1)\right) \leq \mathbb{E}\left(\mathcal{F}(T(R))\right) \leq \frac{1}{2} \tag{5.34}$$

A similar inequality can be derived for the expectation of $T(R)$. If we want to give a more precise evaluation of these moments we first observe by the inverse transformation method (Ross, 1990) and $Z \sim par(2)$ that

$$Z \stackrel{d}{=} U^{-1} - 1$$

with U a standard uniform random variable. Hence, if it is easy to generate a realization of the random variable R, we can always use Monte Carlo simulation to give an more accurate estimate of the expression (see Lemma 5.8)

$$\mathbb{E}\left(\frac{\exp(-R(Z+1))(Z+1)}{(\ln(Z))^2+\pi^2}\right).$$

If we are interested in higher integer moments $\mathbb{E}\left(\mathcal{F}(T(R))^{k+1}\right)$ for some $k \in \mathbb{N}$ we observe that

$$\mathbb{E}\left(\mathcal{F}(T(R))^{k+1}\right) = (k+1)\int_0^1 t^k \mathbb{P}\{\mathcal{F}(T(R)) > t\}dt. \tag{5.35}$$

By the same technique as in Lemma 5.8 we obtain from relation (5.35) that

$$\mathbb{E}\left(\mathcal{F}(T(R))^{k+1}\right) = \frac{1}{k+2} - \pi^{-1}\mathbb{E}\left(\exp(-R(Z+1))ZM_k(-\ln(Z))\right). \tag{5.36}$$

Applying the recurrence relation in Lemma 5.7 we can again use Monte Carlo simulation to estimate the integral in relation (5.36). In the next section we will apply our findings on $\mathcal{F}(T(R))$ and $\lfloor T(R) \rfloor$ to justify the first proposed approximation.

5.3 Approximating the CDF of a randomized hitting time

Since in the next chapter dynamic programming is used to construct an optimal policy and for such a procedure it is too time-consuming to evaluate the CDF of the randomized hitting time $T_{v,\lambda}(R)$ exactly (e.g. via numerical integration), we derive in this section two simple approximations of this CDF. By relation (5.7) it is sufficient to give an approximation of the CDF H_R of the hitting time $T(R)$ listed in relation (5.10). To justify our first approximation we observe by relation (5.12) that

$$\begin{aligned} H_R(t) &= \mathbb{P}\{T(R) \le t\} \\ &= \mathbb{P}\{T(R) - \mathcal{F}(T(R)) \le t - \mathcal{F}(T(R))\} \\ &= \mathbb{P}\{\lfloor T(R) \rfloor \le t - \mathcal{F}(T(R))\} \end{aligned} \tag{5.37}$$

for every $t > 0$. It is shown in Steutel and Thiemann (1990) that

$$cov(T(r), \mathcal{F}(T(r))) = O(\exp(-r)) \tag{5.38}$$

and this implies for every nonnegative random variable R independent of the standard gamma process that

$$cov(T(R), \mathcal{F}(T(R)) = O(\mathbb{E}(\exp(-R))). \tag{5.39}$$

Hence for $\mathbb{E}(\exp(-R))$ small the random variables $T(R)$ and $\mathcal{F}(T(R))$ are practically uncorrelated. Also by the analysis in the previous section we have shown for $\mathbb{E}(\exp(-R))$ small that the random variable $\mathcal{F}(T(R))$ has approximately a uniform distribution on $(0,1)$. Although uncorrelated random variables are in general not independent (equivalence only holds for normal distributed random variables), it seems by the above observations reasonable (for the purpose to obtain easy expressions) to introduce the following approximation assumption. A similar approximation assumption is also introduced in Steutel and Thiemann (1990) for a different purpose.

Approximation assumption 1. The random variables $T(R)$ and $\mathcal{F}(T(R))$ are independent and $\mathcal{F}(T(R))$ is uniformly distributed on $(0,1)$.

Before investigating the consequence of approximation assumption 1 we list the following result.

Lemma 5.9 *If the nonnegative random variable Y is independent of $\mathcal{F}(Y)$, then $\lfloor Y \rfloor$ is also independent of $\mathcal{F}(Y)$.*

Proof. Since $\lfloor Y \rfloor = \sum_{n=0}^{\infty} n 1_{\{n \le Y < n+1\}}$ and hence $\lfloor Y \rfloor$ is a function of the random variable Y the desired result follows. □

If approximation assumption 1 holds we obtain by relation (5.37) and Lemma 5.9 that

$$H_R(t) \approx \mathbb{P}\{\lfloor T(R) \rfloor \le t - U\} \tag{5.40}$$

with U a uniformly distributed random variable independent of $\lfloor T(R) \rfloor$. Introducing now

$$\mathbb{P}\{\lfloor T(R) \rfloor \le -1\} := 0$$

and using the independence of U and $\lfloor T(R) \rfloor$ it follows for every $t > 0$ that

$$\mathbb{P}\{\lfloor T(R) \rfloor \le t - U\} = \mathcal{F}(t))\mathbb{P}\{\lfloor T(R) \rfloor \le \lfloor t \rfloor\} + (1 - \mathcal{F}(t))\mathbb{P}\{\lfloor T(R) \rfloor \le \lfloor t \rfloor - 1\}. \tag{5.41}$$

Also by the continuity of the CDF H_R we obtain for every $t > 0$ that

$$\mathbb{P}\{\lfloor T(R) \rfloor \le \lfloor t \rfloor\} = \mathbb{P}\{T(R) < \lfloor t \rfloor + 1\} = H_R(\lfloor t \rfloor + 1).$$

This implies by relation (5.41)

$$\mathbb{P}\{\lfloor T(R) \rfloor \le t - U\} = \mathcal{F}(t) H_R(\lfloor t \rfloor + 1) + (1 - \mathcal{F}(t)) H_R(\lfloor t \rfloor)$$

and so by relation (5.40) we obtain

$$H_R(t) \approx \mathcal{F}(t) H_R(\lfloor t \rfloor + 1) + (1 - \mathcal{F}(t)) H_R(\lfloor t \rfloor). \tag{5.42}$$

This shows under approximation assumption 1 that the CDF H_R seems to be well approximated by a linear interpolation of H_R at the integer points.

To evaluate the continuous CDF H_R on its integer points we know by relation (5.14) that

$$\begin{aligned} H_R(n+1) - H_R(n) &= \mathbb{P}\{n < T(R) \leq n+1\} \\ &= \mathbb{P}\{\lfloor T(R) \rfloor = n\} \\ &= \tfrac{1}{n!}\mathbb{E}(R^n \exp(-R)) \end{aligned} \tag{5.43}$$

for every $n \in \mathbb{N} \cup \{0\}$. Alternatively, if $\widehat{G}_R(\tau) := \mathbb{E}(\exp(-\tau R))$ is the probability Laplace Stieltjes transform (pLSt) of the CDF G_R, then relation (5.43) is the same as

$$H_R(n+1) - H_R(n) = \frac{(-1)^n}{n!}\widehat{G}_R^{(n)}(1) \tag{5.44}$$

for every $n \in \mathbb{N} \cup \{0\}$ with $\widehat{G}_R^{(n)}, n \in \mathbb{N}$ denoting the nth derivative of $\widehat{G}_R$ and $\widehat{G}_R^{(0)} := \widehat{G}_R$. If the derivatives of $\widehat{G}_R$ are elementary functions we can directly apply relation (5.44) as shown in the next example for R uniformly distributed on $(a, a+b)$. For convenience introduce the sequence p_n, $n \in \mathbb{N} \cup \{0\}$ given by

$$p_n := H_R(n+1) - H_R(n) = \mathbb{P}\{\lfloor T(R) \rfloor = n\}. \tag{5.45}$$

Example 5.10 *If the random variable R is uniformly distributed on $(a, a+b)$ with $a \geq 0$ and $b > 0$ or $R \stackrel{d}{=} a + bU$ with U a standard uniformly distributed random variable we obtain*

$$\widehat{G}_R(\tau) = \mathbb{E}(\exp(-\tau(a+bU))) = b^{-1}(h_a(\tau) - h_{a+b}(\tau)) \tag{5.46}$$

with the functions $h_d : [0, \infty) \to \mathbb{R}$, $d > 0$ given by $h_d(\tau) := \tau^{-1}\exp(-\tau d)$. If we introduce the function $e_n : [0, \infty) \to \mathbb{R}$ given by (Abramowitz and Stegun, 1972)

$$e_n(x) := \sum\nolimits_{j=0}^{n} \frac{x^j}{j!}$$

it is easy to check using Leibniz formula for the differentiation of the product of two functions that for any positive d and τ

$$\begin{aligned} (-1)^n h_d^{(n)}(\tau) &= \exp(-\tau d)\textstyle\sum_{j=0}^{n} \binom{n}{j} d^j \tau^{-(n-j+1)}(n-j)! \\ &= n!\exp(-\tau d)\textstyle\sum_{j=0}^{n} \frac{d^j}{j!}\tau^{-(n-j+1)} \\ &= n!\tau^{-(n+1)}\exp(-\tau d)e_n(\tau d) \end{aligned} \tag{5.47}$$

with $h_d^{(n)}(\tau)$ denoting the nth derivative of the function h_d evaluated in τ. This implies

$$(-1)^n h_d^{(n)}(1) = n!\exp(-d)e_n(d). \tag{5.48}$$

Hence by relations (5.44), (5.45), (5.46) and (5.48) we obtain for every $n \in \mathbb{N} \cup \{0\}$

$$\begin{aligned} p_n &= b^{-1}\left(\exp(-a)e_n(a) - \exp(-(a+b))e_n(a+b)\right) \\ &= b^{-1}(\mathbb{P}\{N(a) \leq n\} - \mathbb{P}\{N(a+b) \leq n\}) \\ &= b^{-1}\mathbb{P}\{N(a) \leq n, N(a+b) > n\} \end{aligned} \tag{5.49}$$

with N *a Poisson process with arrival rate* 1. *For* $n = 0$ *this reduces to*

$$\begin{aligned} p_0 &= b^{-1}\mathbb{P}\{N(a) = 0, N(a+b) - N(a) > 0\} \\ &= b^{-1}\mathbb{P}\{N(a) = 0\}\mathbb{P}\{N(b) > 0\} \\ &= \exp(-a)b^{-1}\left(1 - \exp(-b)\right). \end{aligned}$$

Using relation (5.49) we also obtain for every $n \in \mathbb{N} \cup \{0\}$ *the recurrence relation*

$$p_{n+1} = p_n + b^{-1}\left(\frac{\exp(-a)a^{n+1}}{(n+1)!} - \frac{\exp(-(a+b))(a+b)^{n+1}}{(n+1)!}\right). \tag{5.50}$$

If we consider the special case $a = 0$ *it follows by relation (5.49) for every* $n \in \mathbb{N} \cup \{0\}$ *that*

$$p_n = b^{-1}\mathbb{P}\{N(b) > n\}$$

Hence the recurrence relation in (5.50) reduces to

$$p_{n+1} = p_n - \frac{b^n}{(n+1)!}\exp(-b) \tag{5.51}$$

with $p_0 = b^{-1}(1 - \exp(-b))$.

In some cases we have to use a numerical procedure to evaluate $\widehat{G}_R(1)$. However, as shown by the following example, it is still possible to write down a recurrence relation for the probabilities p_n.

Example 5.11 *If the random variable* $R \stackrel{d}{=} bU^a$ *for some* $b, a > 0$ *and* U *standard uniformly distributed then clearly the domain of* R *is* $(0, b)$ *and*

$$\mathbb{P}\{R \leq x\} = \mathbb{P}\{U \leq (b^{-1}x)^{a^{-1}}\} = (b^{-1}x)^{a^{-1}}$$

for every $0 \leq x \leq b$. *By relation (5.43) we obtain*

$$\begin{aligned} p_n &= \tfrac{b^n}{n!}\mathbb{E}(U^{an}\exp(-bU^a)) \\ &= \tfrac{b^n}{n!}\int_0^1 x^{an}\exp(-bx^a)dx \\ &= \tfrac{a^{-1}b^n}{n!}\int_0^1 y^{n-1+a^{-1}}\exp(-by)dy. \end{aligned} \tag{5.52}$$

Applying Tonelli's theorem it follows for every $\zeta > 0$

$$\begin{aligned}\int_0^1 y^\zeta \exp(-by)dy &= \zeta \int_0^1 \int_0^y x^{\zeta-1}dx \exp(-by)dy \\ &= \zeta \int_0^1 \int_x^1 \exp(-by)dy x^{\zeta-1}dx \\ &= b^{-1}\zeta \int_0^1 \exp(-bx)x^{\zeta-1}dx - b^{-1}\exp(-b).\end{aligned}$$

By this recurrence relation and relation (5.52) it is easy to see for every $n \in \mathbb{N} \cup \{0\}$ *that*

$$p_{n+1} = \frac{n + a^{-1}}{n+1} p_n - \frac{a^{-1}b^n}{(n+1)!}\exp(-b) \tag{5.53}$$

and so for $a = 1$ *we recover relation (5.51). To compute the values* p_n *we first need to compute*

$$p_0 = \mathbb{E}\left(\exp(-bU^a)\right)$$

and this can be done by some numerical integration method (see e.g. Stummel and Hainer, 1980).

A generalization of a uniformly distributed random variable R is given in the following example.

Example 5.12 *Let* G_R *be a concave CDF. It is well-known (Steutel and van Harn, 2004) that a CDF* G_R *of a nonnegative random variable* R *is concave on* $(0, \infty)$ *if and only if* $R \stackrel{d}{=} UY$ *with* U *a standard uniformly distributed random variable,* Y *a nonnegative random variable and* U *and* Y *independent. Using this representation we obtain for every* $\tau > 0$ *that*

$$\widehat{G}_R(\tau) = \int_0^\infty \mathbb{E}\left(\exp(-\tau bU)\right) dG_Y(b)$$

and so by relations (5.44), (5.45) and (5.50) (take $a = 0$*) it follows* $p_0 = \mathbb{E}(\exp(-UY))$ *and*

$$p_{n+1} = p_n - \frac{1}{(n+1)!}\mathbb{E}(Y^n \exp(-Y)) \tag{5.54}$$

for every $n \in \mathbb{N} \cup \{0\}$*. If we introduce the sequence*

$$r_n := \mathbb{P}\{\lfloor T(Y) \rfloor = n\}$$

we obtain by relations (5.14) and (5.54) that

$$p_{n+1} = p_n - \frac{1}{n+1} r_n.$$

This shows that one can evaluate $\mathbb{P}\{\lfloor T(UY) \rfloor = n\}$ *once it is possible to compute the probabilities* $\mathbb{P}\{\lfloor T(Y) \rfloor = n\}$*.*

If $R \sim \text{gamma}(\beta, \rho)$ we obtain by relation (5.44) and (5.45) the recurrence relation

$$p_{n+1} = \frac{\beta + n}{(n+1)(\rho+1)} p_n, \ n \in \mathbb{N} \cup \{0\}, \tag{5.55}$$

with starting value $p_0 = (\frac{\rho}{1+\rho})^\beta$. In the next example we discuss the case of R having an infinitely divisible distribution thus generalizing R having a gamma distribution. A CDF G_R on $(0, \infty)$ is called infinitely divisible if for every $n \in \mathbb{N}$ there exists a sequence of independent, identically distributed and nonnegative random variables R_i, $1 \leq i \leq n$ such that

$$R \stackrel{d}{=} R_1 + \ldots + R_n$$

Examples of infinitely divisible CDF's on $(0, \infty)$ are discussed in Steutel and van Harn (2004) and include gamma distributions and scale mixtures of gamma distributions with shape parameter $\alpha \leq 2$. It is also shown in Steutel and van Harn (2004) that all infinitely divisible CDF's on $(0, \infty)$ can be obtained as weak limits of compound-Poisson distributions.

Example 5.13 *Let the CDF G_R satisfying $G_R(0) = 0$ be infinitely divisible. Necessarily the non-degenerate random variable R has unbounded support (cf. Steutel and van Harn, 2004) and so this excludes nonnegative random variables R with bounded support. Also by Theorem* 4.10 *on page* 95 *and Theorem* 4.14 *on page* 97 *of Steutel and van Harn (2004) it follows that G_R satisfying $G_R(0) = 0$ is an infinitely divisible CDF if and only if there exists some (Laplace-Stieltjes transform) LSt-able function K (the so-called canonical function) satisfying $\int_{(0,\infty)} x^{-1} dK(x) = \infty$ such that*

$$\int_0^x r dG_R(r) = (G_R * K)(x) \tag{5.56}$$

for every $x \geq 0$ with $*$ the well-known convolution operator. A function K is called a LSt-able function (Steutel and van Harn, 2004) if K is right continuous and non-decreasing with $K(x) = 0$ for $x < 0$ and

$$\widehat{K}(\tau) := \int_{0-}^{\infty} \exp(-\tau x) dK(x)$$

is finite for every $\tau > 0$. By Tonelli's theorem and the binomial theorem applied to $(u+v)^{n-1}$ we obtain for $n \in \mathbb{N}$ using relations (5.43), (5.45) and (5.56) that

$$\begin{aligned} n!p_n &= \int_0^\infty \exp(-r) r^n dG_R(r) \\ &= \int_0^\infty \exp(-r) r^{n-1} d(G_R * K)(r) \\ &= \int_{\mathbb{R}_+^2} \exp(-(u+v))(u+v)^{n-1} dG_R(u) dK(v) \\ &= \sum_{j=0}^{n-1} \binom{n-1}{j} \int_0^\infty \exp(-u) u^j dG_R(u) \int_0^\infty \exp(-v) v^{n-1-j} dK(v) \\ &= (n-1)! \sum_{j=0}^{n-1} \tfrac{1}{(n-1-j)!} p_j \int_0^\infty \exp(-v) v^{n-1-j} dK(v). \end{aligned} \tag{5.57}$$

Introducing the constants

$$r_k := \frac{1}{k!} \int_0^\infty x^k \exp(-x) dK(x), \ k \in \mathbb{N} \cup \{0\}$$

this implies by relation (5.57) and (5.43) that

$$p_0 = \widehat{G}_R(1) \text{ and } p_n = \frac{1}{n} \sum\nolimits_{j=0}^{n-1} p_j r_{n-1-j}, \ n \in \mathbb{N}. \tag{5.58}$$

From relation (5.56) it is easy to see (see also Steutel and van Harn, 2004) that

$$\widehat{K}(\tau) = \frac{-\widehat{G}_R^{(1)}(\tau)}{\widehat{G}_R(\tau)}$$

and so by relation (5.58) one can evaluate p_n, $n \in \mathbb{N} \cup \{0\}$ if it is possible to compute the constants r_k, $k \in \mathbb{N}$. Examples are given by Gamma distributions with shape parameter β and scale parameter ρ. For this case the canonical function K has derivative $k(x) = \beta \exp(-\rho x)$ and so $r_n = \beta(\rho+1)^{-(n+1)}$, $n \in \mathbb{N} \cup \{0\}$. This yields $p_n = \frac{\beta}{n} \sum_{j=0}^{n-1} p_j (\rho + 1)^{-(n-j)}$, $n \in \mathbb{N}$, with $p_0 = (\frac{\rho}{1+\rho})^\beta$. This actually defines the same sequence as relation (5.55) does.

Finally, if it is not possible to derive elementary expressions for the nth derivative of the pLSt $\widehat{G}_R$ we observe by relation (5.14) that the generating function (gP) P of the sequence p_n, $n \in \mathbb{N} \cup \{0\}$ is given by

$$P(z) := \mathbb{E}\left(z^{\lfloor T(R) \rfloor}\right) = \int_0^\infty \exp(-r) \sum\nolimits_{n=0}^\infty \frac{(zr)^n}{n!} dG_R(r) = \widehat{G}_R(1-z) \tag{5.59}$$

for every $z \in \mathbb{C}$ with $|z| \leq 1$. Hence we can apply the (numerical) FFT method (Abate and Whitt, 1992) to evaluate the sequence p_n, $n \in \mathbb{N}$.

A second approximation of the CDF H_R is motivated by the following idea. Replace (see Figure 5.1) each sample path of the standard gamma process by a piecewise linear sample path coinciding with the original sample path at integer points and consider this stochastic process X_a instead of the original standard gamma process X. Clearly by construction

$$X_a(n) \overset{a.s.}{=} X(n)$$

and

$$X_a(n+s) = X(n) + s(X(n+1) - X(n))$$

for every $n \in \mathbb{N} \cup \{0\}$ and $0 < s < 1$. Moreover, it follows that $\mathbb{E}(X_a(t)) = \mathbb{E}(X(t))$ for every $t > 0$ and so this new stochastic process has the same expectation at each point as the original one. Alternatively by Lemma 5.5 we know for every non-integer $t > 0$ that

$$X(t) \overset{d}{=} X(\lfloor t \rfloor) + V_{\mathcal{F}(t)} Y \tag{5.60}$$

with $V_{\mathcal{F}(t)} \sim \text{beta}\,(\mathcal{F}(t), 1-\mathcal{F}(t))$, $Y \sim \text{gamma}(1,1)$, $X(\lfloor t \rfloor) \sim \text{gamma}(\lfloor t \rfloor, 1)$ independent random variables. Replacing now in relation (5.60) the random variable $V_{\mathcal{F}(t)}$ by its expectation $\mathcal{F}(t)$ we obtain the proposed stochastic process $X_a = \{X_a(t) : t \geq 0\}$. We now introduce the following approximation assumption.

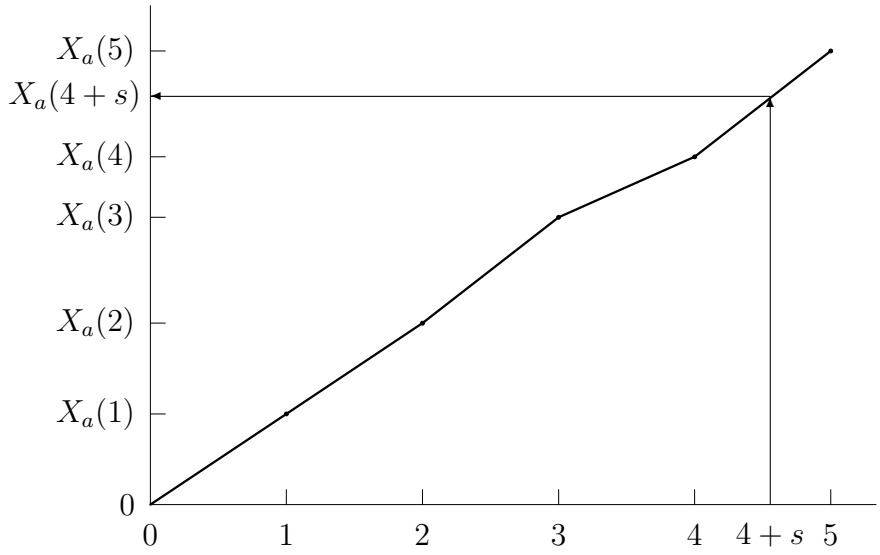

Figure 5.1: Piecewise linear sample path of the process X_a.

Approximation assumption 2. $H_R(t) = \mathbb{P}\{X(t) > R\} \approx \mathbb{P}\{X_a(t) > R\}$.

To evaluate the probability $\mathbb{P}\{X_a(t) > R\}$ we first need the following result.

Lemma 5.14 *For every non-integer $t > 1$ and $x > 0$*

$$(1 - \mathcal{F}(t))\mathbb{P}\{X_a(t) \leq x\} = \mathbb{P}\{X_a(\lfloor t \rfloor) \leq x\} - \mathcal{F}(t)\mathbb{P}\{X_a(t-1) \leq x\}.$$

Proof. If $1 < t < 2$ then $X_a(t) \stackrel{d}{=} X(1) + \mathcal{F}(t)Y_1$ with the random variables $X(1)$ and Y_1 independent and both exponentially distributed with parameter 1. After some calculations this implies for every $x > 0$ that

$$\begin{aligned}(1 - \mathcal{F}(t))\mathbb{P}\{X_a(t) \leq x\} &= (1 - \mathcal{F}(t))\mathbb{P}\{X(1) + \mathcal{F}(t)Y_1 \leq x\} \\ &= \mathbb{P}\{X(1) \leq x\} - \mathcal{F}(t)\mathbb{P}\{\mathcal{F}(t)Y_1 \leq x\}\end{aligned} \tag{5.61}$$

Since

$$X_a(t-1) \stackrel{d}{=} \mathcal{F}(t)Y_1 \text{ and } X(1) = X_a(1)$$

we have verified the desired formula for $0 < t < 1$. To verify the formula for arbitrary non-integer $t > 2$ it follows by the definition of the process X_a that

$$X_a(t) = X_a(\lfloor t \rfloor) + \mathcal{F}(t)Y_1 \stackrel{d}{=} X_a(\lfloor t \rfloor - 1) + X(1) + \mathcal{F}(t)Y_1 \tag{5.62}$$

with the random variables $X(1)$, Y_1 and $X_a(\lfloor t \rfloor - 1)$ independent and $X(1)$, Y_1 exponentially distributed with parameter 1. This implies by relations (5.62) and (5.61) that

$$\begin{aligned}(1 - \mathcal{F}(t))\mathbb{P}\{X_a(t) \leq x\} &= (1 - \mathcal{F}(t))\mathbb{P}\{X_a(\lfloor t \rfloor - 1) + X(1) + \mathcal{F}(t)Y_1 \leq x\} \\ &= \mathbb{P}\{X_a(\lfloor t \rfloor - 1) + X(1) \leq x\} - \mathcal{F}(t)\mathbb{P}\{X_a(\lfloor t \rfloor - 1) + \mathcal{F}(t)Y_1 \leq x\}.\end{aligned}$$

Since $X_a(\lfloor t \rfloor) \stackrel{d}{=} X_a(\lfloor t \rfloor - 1) + X(1)$ and $X_a(t-1) = X_a(\lfloor t \rfloor - 1) + \mathcal{F}(t)Y_1$, the desired result follows. □

By Lemma 5.14 it follows for R independent of the standard gamma process X and hence independent of X_a that

$$(1 - \mathcal{F}(t))\mathbb{P}\{X_a(t) \leq R\} = \mathbb{P}\{X_a(\lfloor t \rfloor) \leq R\} - \mathcal{F}(t)\mathbb{P}\{X_a(t-1) \leq R\}. \tag{5.63}$$

This implies the recurrence relation

$$(1 - \mathcal{F}(t))\mathbb{P}\{X_a(t) > R\} = \mathbb{P}\{X_a(\lfloor t \rfloor) > R\} - \mathcal{F}(t)\mathbb{P}\{X_a(t-1) > R\}. \tag{5.64}$$

To compute in relation (5.63) the probability $\mathbb{P}\{X_a(\lfloor t \rfloor) \leq R\}$ we observe by the continuity of H_R and $X_a(\lfloor t \rfloor) = X(t)$ that

$$\mathbb{P}\{X_a(\lfloor t \rfloor) > R\} = \mathbb{P}\{T(R) \leq \lfloor t \rfloor\} = \mathbb{P}\{\lfloor T(R) \rfloor \leq \lfloor t \rfloor - 1\}$$

and this implies by relations (5.43) and (5.45) that

$$\mathbb{P}\{X(\lfloor t \rfloor) > R\} = \sum\nolimits_{j=0}^{\lfloor t \rfloor - 1} \frac{\mathbb{E}(R^j \exp(-R))}{j!} = \sum\nolimits_{j=0}^{\lfloor t \rfloor - 1} p_j \tag{5.65}$$

for $t > 1$. Finally, for $0 < t < 1$, we obtain using $X_a(t) \stackrel{d}{=} \mathcal{F}(t) Y_1$ with $Y_1 \sim \text{gamma}(1,1)$ that

$$\mathbb{P}\{X_a(t) > R\} = \mathbb{E}(\exp(-\mathcal{F}(t)^{-1} R)) = \widehat{G}_R(\mathcal{F}(t)^{-1}). \tag{5.66}$$

By relation (5.64) up to (5.66) we can compute recursively the value $\mathbb{P}\{X_a(t) > R\}$ in case the pLSt of the random variable R is an elementary expression. In Appendix 5.A an algorithm for computing $\mathbb{P}\{X_a(t) > R\}$ for $t > 0$ is given. In the next section we consider some special cases for which the CDF of $T(R)$ has an elementary expression.

5.4 On the CDF of the randomized hitting time for some special cases

In this section we consider some special cases for which one can give an analytical and/or a simpler probabilistic interpretation of the CDF H_R. We start with the simplest case of R having a degenerate distribution at $r > 0$, i.e. the threshold R is not random but deterministic.

Example 5.15 *If R has a degenerate distribution at $r > 0$, then*

$$H_R(t) = \mathbb{P}\{X(t) \geq r\} = \frac{\Gamma(t, r)}{\Gamma(t)}, \tag{5.67}$$

where $\Gamma(a, x) = \int_{z=x}^{\infty} z^{a-1} \exp(-z) \mathrm{d}z$ is the incomplete gamma function for $a \geq 0$ and $x > 0$. It follows that

$$\mathbb{P}\{T_{v,\lambda}(R) \leq t\} = \frac{\Gamma(v(t), r\lambda)}{\Gamma(v(t))}. \tag{5.68}$$

Another simple case occurs when R has an exponential distribution.

Example 5.16 *If the random variable R has an exponential CDF with scale parameter $\rho > 0$ given by $G_R(r) = 1 - \exp(-\rho r)$, $r > 0$ then by relation (5.11)*

$$H_R(t) = 1 - \mathbb{E}\left(\exp(-\rho X(t))\right) = 1 - (1+\rho)^{-t}. \tag{5.69}$$

This implies

$$T(R) \stackrel{d}{=} \frac{Y}{\ln(1+\rho)}. \tag{5.70}$$

with Y exponentially distributed with parameter 1. Since the random variable λR has an exponential CDF with scale parameter $\rho\lambda^{-1}$ it follows by Lemma 5.2 and relation (5.69) that

$$T_{v,\lambda}(R) \stackrel{d}{=} v^{\leftarrow}\left(\frac{Y}{\ln(1+\rho\lambda^{-1})}\right). \tag{5.71}$$

and this shows

$$\mathbb{P}\{T_{v,\lambda}(R) \leq t\} = 1 - (1+\rho\lambda^{-1})^{-v(t)}.$$

For the power function $v(t) = t^q$, $q > 0$ we obtain that $v^{\leftarrow}(u) = u^{q^{-1}}$ and so by relation (5.71) this yields

$$T_{v,\lambda}(R) \stackrel{d}{=} \frac{Y^{q^{-1}}}{(\ln(1+\rho\lambda^{-1}))^{q^{-1}}}.$$

Since Y is exponentially distributed with parameter 1 and hence $Y^{q^{-1}}$ is Weibull distributed with shape parameter q and scale parameter 1 (Barlow and Proschan, 1975) we obtain that $T_{v,\lambda}(R)$ is a Weibull distributed random variable with shape parameter q and scale parameter $(\log(1+\rho\lambda^{-1}))^{q^{-1}}$.

A generalization of the exponential distribution is given in the following example.

Example 5.17 *If the positive random variable R has density function g_R on $(0,\infty)$ belonging to the class CM of completely monotone densities g, i.e. g is nonnegative and $(-1)^n g^{(n)}(x) \geq 0$ for $x > 0$ and $n \in \mathbb{N}$ with $g^{(n)}(\cdot)$ denoting the nth derivative of g, then by Bernstein's theorem (Feller, 1971; Steutel and van Harn, 2004) we obtain*

$$g_R(r) = \int_0^\infty y\exp(-yr)dG(y),\ r > 0$$

for some CDF G on $(0,\infty)$. This shows $g_R \in CM$ if and only if $R \stackrel{d}{=} ZY$ with Y exponentially distributed with parameter 1, Z a nonnegative random variable and Z independent of Y. By relation (5.70) and zY has an exponential distribution with scale parameter z^{-1} this yields

$$T(R) \stackrel{d}{=} \frac{Y}{\ln(1+Z^{-1})} \tag{5.72}$$

and so for R having a mixture of exponential distributions also the CDF H_R has a mixture of exponential distributions. In general it is not possible to give a nice analytical expression for the CDF H_R unless the CDF G_R has a finite mixture of exponential distributions given by

$$G_R(r) = \sum\nolimits_{k=1}^{q} p_k(1-\exp(-\lambda_k r))$$

with $0 < \lambda_1 < \ldots < \lambda_q$ and $p_i > 0$, $\sum_{k=1}^{q} p_k = 1$. This means that the discrete distribution of Z is given by

$$\mathbb{P}\{Z = \lambda_k^{-1}\} = p_k.$$

and by relation (5.72) we obtain

$$H_R(t) = \sum_{k=1}^{q} p_k(1-(1+\lambda_k)^{-t}) = 1 - \sum_{k=1}^{q} p_k(1+\lambda_k)^{-t}. \tag{5.73}$$

Similarly it follows by Lemma 5.2, $\lambda R \stackrel{d}{=} \lambda ZY$ *and relation (5.72) that*

$$T_{v,\lambda}(R) \stackrel{d}{=} v^{\leftarrow}\left(\frac{Y}{\ln(1+(\lambda Z)^{-1})}\right) \tag{5.74}$$

and this implies for $v(t) = t^q$, $q > 0$ *that*

$$T_{v,\lambda}(R) \stackrel{d}{=} \frac{Y^{q^{-1}}}{\ln(1+(\lambda Z)^{-1})^{q^{-1}}}.$$

As before for R *having a finite mixture of exponential distributions it is easy to check that the CDF* H_R *has a nice analytical expression given by a finite mixture of Weibull distributions. Moreover, if it is possible to generate a realization of the random variable* Z *it is easy by relation (5.74) to estimate the CDF of* $T_{v,\lambda}(R)$ *by Monte Carlo simulation.*

If the positive random variable R has a gamma distribution with shape parameter $0 < \beta < 1$ and scale parameter 1 it follows by Lemma 5.5 that R has a completely monotone density. In particular one can show the following.

Example 5.18 *If the random variable* R *has a gamma distribution with shape parameter* $0 < \beta < 1$ *and scale parameter* 1*, then by Lemma 5.5 we obtain that*

$$R \stackrel{d}{=} ZY$$

with Y *and* Z *independent,* $Y \sim$ *gamma*$(1,1)$ *and* $Z \sim$ *beta*$(\beta, 1-\beta)$*. Hence by relation (5.72) we obtain*

$$T(R) \stackrel{d}{=} \frac{Y}{\ln(1+Z^{-1})} \tag{5.75}$$

Since $Z \sim$ *beta*$(\beta, 1-\beta)$ *it is well-known (Hogg and Craig, 1978; Wilks, 1962) that* $Z \stackrel{d}{=} Z_\beta(Z_\beta + Z_{1-\beta})^{-1}$ *with* Z_β, $Z_{1-\beta}$ *independent and* $Z_\beta \sim$ *gamma*$(\beta, 1)$ *and* $Z_{1-\beta} \sim$ *gamma*$(1-\beta, 1)$*. This implies by relation (5.75) that*

$$T(R) \stackrel{d}{=} \frac{Y}{\ln(2+Z_{1-\beta}Z_\beta^{-1})}.$$

By Lemma 5.2 it follows similarly

$$T_{v,\lambda}(R) \stackrel{d}{=} v^{\leftarrow}\left(\frac{Y}{\ln(1+\lambda+Z_{1-\beta}Z_\beta^{-1}) - \ln(\lambda)}\right)$$

and so for $v(t) = t^q$, $q > 0$ *we obtain*

$$T_{v,\lambda}(R) \stackrel{d}{=} \frac{Y^{q^{-1}}}{\left(\ln(1+\lambda+Z_{1-\beta}Z_\beta^{-1}) - \ln(\lambda)\right)^{q^{-1}}}.$$

Another generalization of the exponential CDF is given by the following.

Example 5.19 *If the positive random variable R follows a gamma distribution with shape parameter $m \in \mathbb{N}$ and scale parameter $\rho > 0$ we obtain*

$$G_R(r) = 1 - \exp(-\rho r) \sum\nolimits_{j=0}^{m-1} \frac{(\rho r)^j}{j!}.$$

Hence by relation (5.11) it follows that

$$H_R(t) = 1 - \sum\nolimits_{j=0}^{m-1} \frac{\rho^j}{j!} \mathbb{E}(X(t)^j \exp(-\rho X(t))). \tag{5.76}$$

To evaluate the expressions in relation (5.76) we observe using $X(t) \sim$ gamma$(t, 1)$ that

$$\mathbb{E}(X(t)^j \exp(-\rho X(t))) = \frac{1}{\Gamma(t)} \int_0^\infty \exp(-(1+\rho)x) x^{j+t-1} dx \tag{5.77}$$

It is now easy to show by its relation with a gamma distribution with scale parameter $\lambda > 0$ and shape parameter $j + t$ that

$$\frac{1}{j!\Gamma(t)} \int_0^\infty \exp(-\lambda r) r^{j+t-1} dr = \frac{\Gamma(j+t)}{\Gamma(t) j! \lambda^{j+t}} = \binom{t+j-1}{j} \lambda^{-(j+t)} \tag{5.78}$$

with

$$\binom{k}{j} := \begin{cases} \frac{\prod_{p=0}^{j-1}(k-p)}{j!} & j \in \mathbb{N} \\ 1 & j = 0 \end{cases}, \; k \in \mathbb{R},$$

for every $\lambda > 0$, $t > 0$ and $j \in \mathbb{N} \cup \{0\}$. Hence it follows for every $j \in \{0, ..., m-1\}$

$$\mathbb{E}(X(t)^j \exp(-\rho X(t))) = j! \binom{t+j-1}{j} (1+\rho)^{-(j+t)}$$

and this implies using relation (5.76) that

$$H_R(t) = 1 - (1+\rho)^{-t} \sum\nolimits_{j=0}^{m-1} \binom{t+j-1}{j} \left(\frac{\rho}{1+\rho}\right)^j. \tag{5.79}$$

Moreover, since λR has a gamma distribution with shape parameter $m \in \mathbb{N}$ and scale parameter $\lambda^{-1}\rho$ we obtain by relation (5.79) that

$$H_{\lambda R}(t) = 1 - (\lambda^{-1}\rho + 1)^{-t} \sum\nolimits_{j=0}^{m-1} \binom{t+j-1}{j} \left(\frac{\rho}{\lambda+\rho}\right)^j$$

and applying lemma 5.2 yields

$$\mathbb{P}\{T_{v,\lambda}(R) \leq t\} = 1 - (\lambda^{-1}\rho + 1)^{-t} \sum\nolimits_{j=0}^{m-1} \binom{v(t)+j-1}{j} \left(\frac{\rho}{\lambda+\rho}\right)^j.$$

Distribution	Parameters
Deterministic (det)	$r > 0$
Uniform (unif)	$a, a+b,\ a, b > 0$
Erlang (erlang)	$m \in \mathbb{N},\ \rho > 0$
Mixture of two exponentials (hyp-2)	$p_1, \lambda_1, p_2, \lambda_2 > 0,\ p_1 + p_2 = 1,\ \frac{p_1}{\lambda_1} = \frac{p_2}{\lambda_2}$

Table 5.1: The four distributions for R under consideration.

5.5 Numerical study

This section presents the computational results of the two proposed approximations of the randomized hitting times discussed in Section 5.3. We first focus on the accuracy of the approximations of the CDF of the random variable $T(R)$. Next we also compare the computing time of the approximations and the exact (numerical) evaluation of the CDF of $T_{v,\lambda}(R)$. To serve our purposes we have written a computer program using MATLAB 7.2 on a Pentium III–2 GHz personal computer.

In our first experiments we consider four distributions for the random variable R and these are given in Table 5.1. The discrete mixture of two exponentials in Table 5.1 is also referred to as a hyperexponential-2 with balanced means (Tijms, 2003). For these four distributions analytical expressions for the CDF H_R exist and this enables us to assess the accuracy of the approximations.

The accuracy of the approximations is measured by the maximum absolute difference between the approximative and the true CDF on its entire domain. By relation (5.9) we only have to focus on the accuracy of the used approximation of the CDF $H_{\lambda R}$. To give an approximation of the supnorm error we first evaluate the true CDF $H_{\lambda R}$ and the used approximation for values of t on a grid $\{ih\}_{i=1,2,\ldots,N}$, with $h = 0.005$ and

$$N = \inf\{n \in \mathbb{N} :\ \mathbb{P}\{T(\lambda R) \leq \lceil nh \rceil\} > 1 - 10^{-7}\}.$$

Subsequently, the right-hand side of (5.9) is estimated by the maximum absolute difference between the approximations and the true CDF on this grid, i.e.

$$\max_{i=1,2,\ldots,N} |A_{\lambda R}(ih) - H_{\lambda R}(ih)|,\ h > 0.$$

Since for the first proposed approximation we have to compute the CDF of $T(R)$ at integer points and then interpolate, it is straightforward to obtain the approximation of this CDF at non-integer points. The algorithms presented in Appendix 5.A are implemented to compute the second approximation of H_R. Since it is possible to vary the scale parameter λ of a stationary gamma process in such a way that the first hitting time of R can take a wide range of values we fix the value of the expectation of R to 100. This means for

a degenerate random variable that $r = 100$. Next, we vary the coefficient of variation $CV(R)$ of R, given by $CV(R) = (\mathrm{Var}(R))^{1/2} \,/\, \mathbb{E}(R)$, from 0 ($R$ deterministic) to 1.2 (R mixture of two exponentials) by steps of 0.2. The parameters of the three non-degenerate distributions are now determined by means of a two-moment fit (Tijms, 2003). The coefficient of variation of a uniform random variable is bounded from above by $\sqrt{3}/3$. The discrete mixture of two exponentials with balanced means has only two free parameters and its coefficient of variation is bounded from below by 1. The coefficient of variation of the Erlang distribution can take any value greater than 0.

In Table 5.2 the maximum absolute differences between the approximations and the true CDF of $T(\lambda R)$ are shown for a range of values of λ. The higher the value of λ the higher the accuracy of the approximations. Both methods appear to be quite accurate for $\lambda \geq 0.05$. Since for λ increasing the expectation $\mathbb{E}(\exp(-\lambda R))$ becomes smaller this is to be expected from our theoretical results in Section 5.2. The smaller this expectation the more the fractional part of $T_{\lambda R}$ has a uniform CDF and at the same time the correlation between the fractional and integer part of $T_{\lambda R}$ vanishes. This makes our approximative assumption more accurate. The effect of the coefficient of variation of the random variable R on precision is not unambiguous. In Figure 5.2 the accuracy of the approximations is plotted against λ (on a double-logarithmic scale) for R Erlang distributed with $CV(R) = 1.2$. We observe that the maximum absolute error is smaller than 0.01 for $\lambda > 0.02$. In Figure 5.3 the exact and approximative CDF's of $T(\lambda R)$ are plotted against time t for uniformly distributed R with $CV(R) = 0.2$ and $\lambda = 0.02$. It appears that the left tail of the distribution is not approximated very well.

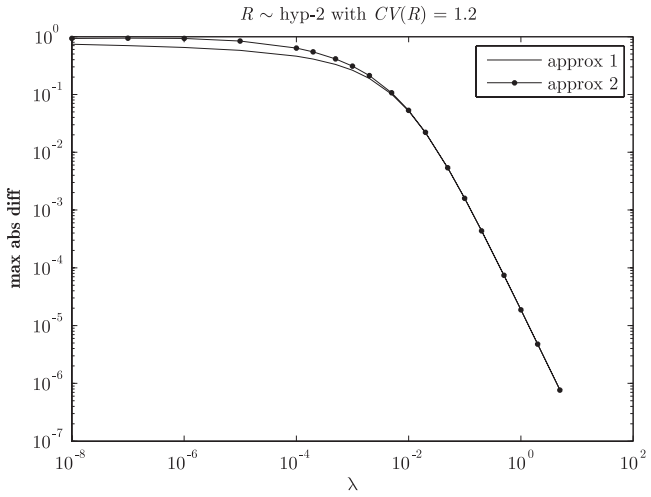

Figure 5.2: Maximum absolute difference between $\mathbb{P}\{T(\lambda R) \leq t\}$ and the two approximations for different values of λ and $R \sim$ hyp-2 with $CV(R) = 1.2$.

					λ		
R	$CV(R)$	Approx	0.0005	0.005	0.05	0.5	5
Deterministic	0	1	2.85E−1	4.32E−2	7.31E−3	6.35E−4	6.14E−5
		2	3.98E−1	5.57E−2	7.83E−3	6.38E−4	6.15E−5
Uniform	0.2	1	2.86E−1	4.24E−2	6.66E−3	2.52E−4	7.97E−6
		2	3.98E−1	5.53E−2	7.12E−3	2.53E−4	7.97E−6
	0.4	1	2.92E−1	3.98E−2	5.72E−3	1.84E−4	5.81E−6
		2	3.98E−1	5.96E−2	6.52E−3	1.86E−4	5.81E−6
Erlang	0.4	1	2.96E−1	4.08E−2	6.33E−3	1.19E−4	1.32E−6
		2	4.05E−1	6.41E−2	6.74E−3	1.20E−4	1.33E−6
	0.6	1	2.90E−1	3.67E−2	6.09E−3	8.43E−5	8.85E−7
		2	3.87E−1	5.88E−2	6.46E−3	8.44E−5	8.87E−7
	0.8	1	2.79E−1	3.45E−2	5.14E−3	1.11E−4	1.21E−6
		2	3.61E−1	5.07E−2	5.32E−3	1.11E−4	1.21E−6
Hyp-2	1	1	3.24E−1	9.01E−2	3.80E−3	4.85E−5	4.99E−7
		2	4.04E−1	9.43E−2	3.80E−3	4.85E−5	4.99E−7
	1.2	1	3.31E−1	1.01E−1	5.35E−3	7.40E−5	7.67E−7
		2	4.13E−1	1.07E−1	5.36E−3	7.40E−5	7.67E−7

Table 5.2: The maximum absolute difference between the approximation and the true CDF for different values of λ and various choices of the random variable R.

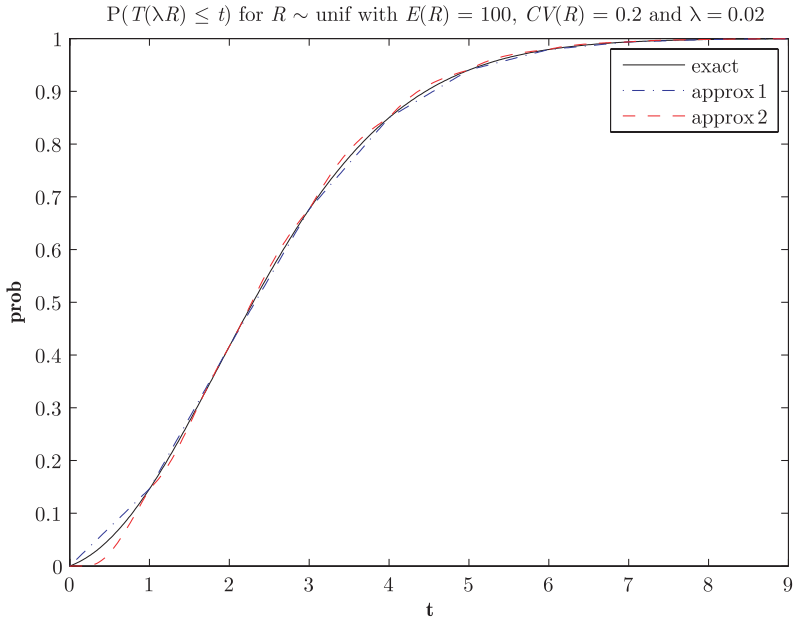

Figure 5.3: Exact and approximative cumulative distribution functions of $T(\lambda R)$, where $R \sim$ unif with $CV(R) = 0.2$ and $\lambda = 0.02$.

We have seen that the scale parameter of the gamma process rules the accuracy of the approximations. However, the computing time of the true CDF of the random variable $T_{v,\lambda}(R)$ is affected by both the shape function and the scale parameter. In order to assess this effect, we focus on non-standard gamma processes $X_{v,\lambda}$ having a shape function v proportional to a power of time, i.e. $v(t) = \nu t^q$ with $\nu > 0$, and scale parameter $\lambda > 0$. We set the value of the power q at 0.5, 1, 2 and 4 and thus cover concave as well as convex shape functions. The mean of the non-stationary gamma process at time t_0 is fixed at $x = \lambda^{-1}\nu t_0^q = 100$ and for t_0 we consider the values 10, 20, 50 and 100. The coefficient of variation of the gamma process at time t_0 is varied from 0.1 to 0.7 by steps of 0.2. For given values of q, t_0 and $c = CV(X_{v,\lambda}(t_0)) = \nu^{-1/2}{t_0}^{-q/2}$, the parameters ν and λ are now determined by a two-moment fit. This yields $\nu = c^{-2}t_0^{-q}$ and $\lambda = c^{-2}x^{-1}$. Observe a high value of c implies a small value of λ. Therefore, the approximation of the CDF of a randomized hitting time associated with non-stationary gamma process is accurate whenever the variability of the gamma process is not too high. For obvious reasons we are only interested in the computing times of good approximations. In the scenarios defined by the experimental settings in Table 5.3 the value of λ ranges from 0.020 ($c = 0.7$) to 1 ($c = 0.1$) and hence by our previous findings the approximations are accurate.

The computing time is defined as the time needed to evaluate or approximate the CDF of $T_{v,\lambda}(R)$ on a grid $\{ih\}_{i=1,2,\ldots,N}$, where $h = 0.02$ and N is the first integer such that the

Settings $X_{v,\lambda}$	Values
$t_0 = \{t > 0 : \mathbb{E}(X_{v,\lambda}(t)) = 100\}$	10, 20, 50, 100 time units
q	0.5, 1, 2, 4
$CV(X_{v,\lambda}(t_0))$	0.1, 0.3, 0.5, 0.7

Table 5.3: Experimental settings.

			q			
R	$CV(R)$	Evaluation	0.5	1	2	4
Deterministic	0	exact	4.14	1.86	1.23	1.04
		approx1	1.10E−2	3.80E−3	2.94E−3	5.99E−3
		approx2	9.63E−2	4.24E−2	5.58E−2	5.68E−2
Uniform	0.2	exact	19.07	8.06	5.20	4.17
		approx1	3.64E−2	2.87E−2	2.75E−2	3.15E−2
		approx2	1.99E−1	1.39E−1	1.46E−1	1.46E−1
Uniform	0.4	exact	24.14	8.89	5.60	4.55
		approx1	4.84E−2	3.69E−2	3.60E−2	3.96E−2
		approx2	2.29E−1	1.40E−1	1.48E−1	1.89E−1

Table 5.4: Computing times for different values of q and different random variables R. Here, $t_0 = 50$ and $CV(X_{v,\lambda}(t_0)) = 0.3$ yielding $\lambda = \frac{1}{9}$.

CDF in the integer point $\lceil v(Nh) \rceil$ exceeds 0.999. We only focus on R deterministic and R uniform, since for these random variables the CDF of $T_{v,\lambda}(R)$ has no nice analytical expression.

In Table 5.4 the computing times of the approximations and the true CDF are given for one particular scenario of Table 5.3. Note that the value of λ is constant and so the hitting times of the (non-stationary) gamma processes are all related to the same random variable $T(\lambda R)$. The results show that the efforts of obtaining both approximations is much less than the effort of computing the true CDF. In all cases the first approximation can be obtained faster than the second approximation. Also, increasing values of q are negative related to the effort of computing the true CDF. This is caused by the fact that the right tail of the hitting time distribution is smaller for larger values of q. Thus for fixed values of $\mathbb{E}(R)$ and $CV(R)$ the CDF is computed in fewer points for q large. This effect is not that apparent for the approximations since their computing times are mainly

affected by the number of *integer* time points in which the CDF needs to be computed and this again depends on the value of λ. The experiments with the other scenarios yield similar results and are available upon request.

5.6 Conclusions

The gamma process plays an important role in maintenance optimization. In particular, the first time at which this process exceeds a random threshold is often used to model the lifetime of structures subject to deterioration. In this article we have investigated in detail the CDF of this random variable. We have first shown that the CDF of a randomized hitting time associated with a non-stationary gamma processes is easily derived from the CDF of a similar hitting time of a standard gamma process. Secondly, we have extended an existing result on the CDF of the fractional part of a randomized hitting time.

Explicit formulas for the CDF of a randomized hitting time have been derived for some special cases. In general however, the evaluation of the CDF of a randomized hitting time for a standard gamma process is time-consuming. Therefore, we have proposed two approximations having a clear probabilistic interpretation. The first approximation comes down to a linear interpolation of the exact hitting time CDF at integer points and it is justified by above-mentioned result on the CDF of the fractional part of the hitting time. We have shown that due to the structure of a standard gamma process it is possible to compute the probability distribution of a randomized hitting time at integer time points. The second approximation is obtained by replacing each sample path of a standard gamma process by a piecewise linear sample path coinciding with the original sample path at integer points.

Numerical experiments show that both approximation formulas are quite accurate when the random threshold is not too small and the variability of the (non-stationary) gamma process is not too high. The second approximation method may be somewhat prohibitive from a numerical point of view. On the other hand, the first approximation is quite efficient and appears to be a good replacement for the exact distribution in time-consuming optimization algorithms.

5.A Algorithm for approximation 2

The algorithm below computes $\mathbb{P}\{X_a(t) > R\}$ for $t > 0$.

Algorithm 5.20 (Forward approximation algorithm)
Input: time $t > 0$. *Output:* $\mathbb{P}\{X_a(t) > R\}$.

(1) For $k = 0$ *to* $\lfloor t \rfloor$
Let cdfint(k) $= \sum_{j=0}^{k-1} p_j$*, where* $\sum_{j=0}^{-1} p_j = 0$ *(cdfint(k)* $= \mathbb{P}\{X_a(k) > R\}$*).*
Next k
Here p_j *is computed using the right-hand side of* (5.44)*. If* $\mathcal{F}(t) = 0$*, then return cdfint(*$\lfloor t \rfloor$*). Otherwise, proceed with step (2).*

(2) Let cdf $= \widehat{G}_R(\mathcal{F}(t)^{-1})$ *(cdf* $= \mathbb{P}\{X_a(\mathcal{F}(t)) > R\}$*).*

(3) For $j = 1$ *to* $\lfloor t \rfloor$
Let cdf $= (cdfint(j) - cdf \cdot \mathcal{F}(t))/(1 - \mathcal{F}(t))$ *(cdf* $= \mathbb{P}\{X_a(t) > R\}$*,* $j < t < j+1$*).*
Next j
Return cdf.

Unfortunately, if the values of cdf and cdfint are close to 1 and $\mathcal{F}(t) > 0.5$, the repeated subtraction in step (3) leads to loss of precision for $t > 1$. These numerical difficulties are circumvented by employing a backward version of algorithm 5.20.

Algorithm 5.21 (Backward approximation algorithm)
Input: time $t > 1$ *with* $\mathcal{F}(t) > 0.5$*. Output:* $\mathbb{P}\{X_a(t) > R\}$*.*
Let M *be a large integer, say* $M = 100$*.*

(1) For $k = 0$ *to* $\lfloor t + M \rfloor$ *+1*
Let cdfint(k) $= \sum_{j=0}^{k-1} p_j$*, where* $\sum_{j=0}^{-1} = 0$ *(cdfint(k)* $= \mathbb{P}\{X_a(k) > R\}$*).*
Next k
Here p_j *is computed using expression* (5.44)*.*

(2) Let cdf=cdfint($\lfloor t + M \rfloor + 1)\mathcal{F}(t)$*+ cdfint(*$\lfloor t + M \rfloor)(1 - \mathcal{F}(t))$*. (cdf here represents the approximate value of the cdf at time* $M + \mathcal{F}(t)$ *according to method* 1*).*

(3) For $j = \lfloor t \rfloor + C$ *to* $\lfloor t \rfloor + 1$ *step* (-1)
Let cdf $= (cdfint(j) - cdf \cdot (1 - \mathcal{F}(t)))/\mathcal{F}(t)$ *(cdf* $= \mathbb{P}\{X_a(t) > R\}$*,* $j < t < j+1$*).*
Next j
Return cdf.

In step (2) we estimate the CDF at time $M + \mathcal{F}(t)$ by linear interpolation at the surrounding integers (approximation method 1). The estimate does not even have to be very accurate, because the backwards algorithm gains precision in every step (as opposed to the forward algorithm).

Finally, note that in order to compute the CDF $\mathbb{P}\{X_a(t) > R\}$ one needs to compute the CDF at time points $\mathcal{F}(t), \mathcal{F}(t)+1, \ldots, t-1$ (forward algorithm) or $M+\mathcal{F}(t), M-1+\mathcal{F}(t), \ldots, t+1$ (backward algorithm). So, if one wants to compute the CDF at equidistant points ih, $i = 1, \ldots, N$, for some $N \in \mathbb{N}$ and $h > 0$, one only has to compute the

CDF at the greatest (smallest) time points with different fractional parts in the forward (backward) algorithm. The other CDF values are obtained for free.

Chapter 6

Modelling and optimizing imperfect maintenance actions with application to coatings on steel structures*

Abstract

Steel structures such as bridges, tanks and pylons are exposed to outdoor weathering conditions. In order to prevent them from corrosion they are protected by an organic coating system. Unfortunately, the coating system itself is also subject to deterioration. Imperfect maintenance actions such as spot repair and repainting can be done to extend the lifetime of the coating. In this chapter we consider the problem of finding the set of actions that minimizes the expected (discounted) maintenance costs over both a finite horizon and an infinite horizon. To this end we model the size of the area affected by corrosion by a non-stationary gamma process. An imperfect maintenance action is to be done as soon as a fixed threshold is exceeded. The direct effect of such an action on the condition of the coating is assumed to be random. On the other hand, due to maintenance the parameters of the gamma deterioration process may also change. It is shown that the optimal maintenance decisions related to this problem are a solution of a continuous-time renewal-type dynamic programming equation. To solve this equation time is discretized and it is verified theoretically that this discretization induces only a small error. Finally, the model is illustrated with a numerical example.

6.1 Introduction

Steel structures such as bridges, tanks and pylons are exposed to outdoor weathering conditions. In order to prevent them from corrosion they are protected by organic coating

*This chapter is based on Nicolai, Frenk, and Dekker (2007b)

systems. Unfortunately, the coating system itself is also subject to deterioration and after some time the steel loses its coating and starts corroding. Maintenance can be done to improve the condition of the coating system and by doing so the lifetime of the steel structure is also extended. Typical maintenance actions for coating systems are (local) spot repair, repainting and replacement. Spot repair consists of only painting the most visible corroded parts, while repainting means that the entire surface of the structure is repainted without removing the corrosion completely. Finally, in a replacement action the old coating and all corrosion is completely removed and a new coating is applied. Since in spot repair and repainting some corrosion is not removed these actions can be seen as imperfect. Obviously the replacement action restores the condition of the coating to new and therefore it is a perfect maintenance action. With respect to cost it is obvious that spot repair is the cheapest action, while replacement is the most expensive.

The aim of this study is to find an optimal strategy for imperfect maintenance of engineering structures, in particular steel structures protected by coatings. To this end, we introduce a deterioration model that includes the effect of imperfect maintenance. The above-defined actions are then employed to form a maintenance strategy and they are the basis of our optimization model. The objective of this model is to minimize the expected maintenance costs over a finite horizon. Clearly this finite horizon is determined by the economic or technical lifetime of the structure. The decision variables are the maintenance actions to be executed.

To model the deterioration process of the coating we use a non-stationary gamma process with state space the size of the coating area affected by corrosion (in e.g. the number of squared meters). As far as the authors know, Abdel-Hameed (1975) was the first to propose the gamma process as a proper model for continuous time-dependent deterioration. The gamma process has increasing sample paths and as such it is a suitable candidate to describe the (monotone) deterioration of engineering structures, see e.g. Çinlar *et al.* (1977), Van Noortwijk and Klatter (1999) and Frangopol *et al.* (2004). In particular, in Heutink *et al.* (2004) and Nicolai *et al.* (2007a) the deterioration of coatings on steel structures is modelled by a non-stationary gamma process. For more examples of the application of gamma processes in maintenance we refer to a recent overview by Van Noortwijk (2007).

Note that the Wiener process or Brownian motion is also often used to describe time-dependent deterioration (see e.g. Doksum and Hóyland (1992) and Whitmore and Schenkelberg (1997)). In contrast to the gamma process, being a jump process with an infinite number of infinitesimal jumps in each finite interval, Brownian motion has continuous sample paths. As such, it is the only so-called Lévy process with this property. On the other hand, the increments of Brownian motion can be negative. In Chapters 2 and 4 we have seen that due to this feature, Brownian motion is *not* a proper model for

deterioration when the variability in deterioration is high. Therefore, we use here the non-stationary gamma process to describe deterioration.

The non-stationary gamma process has also been used to model deterioration in the presence of imperfect maintenance. Van Noortwijk and Frangopol give a mathematical description of the lifetime extending maintenance (LEM) model for engineering structures, which was introduced by Bakker *et al.* (1999). In this model an imperfect maintenance action reduces the amount of deterioration by a fixed amount and afterwards deterioration is again modelled by the same non-stationary gamma process. We shall argue below that the reduction in deterioration may also be random in practice and secondly that the structural parameters of the deterioration process may change after doing an imperfect maintenance action. The present application thus asks for a more complex model.

For steel structures, maintenance actions such as spot repair improve the deterioration rate of the coating only locally, whereas other parts of the surface still deteriorate at the same rate. So, as a whole the deterioration process may increase faster after spot repair than after replacement. The same holds for the repainting action. So, we have to extend the LEM model and therefore we allow for a structural change in the gamma deterioration process (e.g. a time transformation) after maintenance is done.

In practice (imperfect) coating maintenance is often done as soon as the area affected by corrosion exceeds a certain intervention level set by the decision maker. An imperfect maintenance action reduces the size of the affected area by a *random* amount. This random effect occurs since spot repair and repainting do not cover all corrosion as not all may be visible. Observe all corrosion is removed by a replacement. As the improvement in deterioration is modelled by a nonnegative random variable, the time between two maintenance actions is given by the time the gamma process needs to counterbalance this improvement. This time depends on the parameters of the gamma process and the random improvement. With respect to the latter, we consider generally distributed random improvements in deterioration independent of the gamma process.

In the literature several optimization models for deteriorating systems with imperfect maintenance have been proposed. Newby and Barker (2006) extend the general condition-based inspection / replacement model by Newby and Dagg (2004) with imperfect repair, where the effect of imperfect repair is fixed. Deterioration is modelled as a (stationary) Lévy process. A failure is defined as the event in which the process exceeds a fixed failure level. Failure is detected only by (aperiodic) inspections. If the system has failed it is correctively replaced and otherwise a partial repair is carried out. In Newby and Barker (2006) dynamic programming is used to optimize both the expected average and the discounted maintenance costs with respect to the inspection times.

Castanier *et al.* (2003) proposes a similar condition-based maintenance model as Newby and Dagg (2004), but they model stationary deterioration in discrete-time. It

is assumed that the effect of imperfect repair is a random function of the observed deterioration. Partial repair and replacement are done as soon as the observed deterioration exceeds certain thresholds. The expected average cost per unit time is minimized with respect to these thresholds. Meier-Hirmer *et al.* (2005) have applied a version of this model to optimize the maintenance of high-speed railway tracks. The above-mentioned LEM model is an age replacement model with the possibility of partial repair. Finally, Liao *et al.* (2006) also consider systems subject to measurable deterioration, where the effect of partial repair is random. As opposed to the other papers, they consider continuous-monitoring and they optimize the availability of the deteriorating system with respect to a preventive replacement threshold.

In our optimization model maintenance is done as soon as a fixed intervention level is exceeded; no inspections are done. As there are different maintenance actions available for steel structures, the main interest of this paper is in finding the sequence of actions that minimizes the expected maintenance costs over a finite time horizon. This problem can be formulated as a continuous-time renewal-type dynamic programming equation. Time is discretized to solve this equation and it is shown that the solution of the discrete time problem is close in the supnorm to the solution of the original continuous-time problem. This is supported by numerical evidence. Note that although this chapter has a clear focus on optimal imperfect maintenance of coating systems, the described methods should be regarded as generally applicable.

The outline of this chapter is as follows. In Section 6.2 we introduce a deterioration model for structures subject to imperfect maintenance. The associated continuous-time dynamic programming equation describing the optimal maintenance actions is presented and analyzed in Section 6.3. We also do an error analysis for the discretization of this equation. In Section 6.4 we briefly review some techniques presented in Chapter 5 to compute the cumulative distribution function of the time between two maintenance actions. Next, in Section 6.5, we employ these techniques to solve the optimization problem formulated in Section 6.3. In Section 6.6 we draw conclusions.

6.2 Modelling deterioration and maintenance

In this section we present a deterioration model for coating systems on corroding structures undergoing imperfect maintenance actions. The deterioration process of the protective coating is given by a non-stationary gamma process and maintenance is done as soon as the size of the affected area exceeds a given level $\rho > 0$ set by the decision maker. Imperfect maintenance yields a random reduction in this size, bringing it back between 0 and ρ. Next, the deterioration process of the coating is again modelled by a, possibly

different, non-stationary gamma process. In the remainder of this chapter, boldfaced letters are used to denote random variables.

6.2.1 Deterioration model

In this chapter the deterioration process of the coating is represented by a non-stationary gamma process. To introduce the definition of a gamma process we first observe (Steutel and van Harn, 2004) that the density of a gamma distributed random variable with shape parameter $\beta > 0$ and scale parameter $\lambda > 0$ is given by

$$f(x) = \Gamma(\beta)^{-1}\lambda^{\beta}x^{\beta-1}\exp(-\lambda x)1_{(0,\infty)}(x)$$

with

$$\Gamma(\beta) := \int_0^{\infty} x^{\beta-1}\exp(-x)\mathrm{d}x$$

the well-known gamma function. The cumulative distribution function (CDF) of such a random variable is denoted by gamma(β, λ). Also we mean by $X \stackrel{d}{=} Y$ that the random variables X and Y have the same CDF and by $X \sim F$ that the random variable X has CDF F.

Definition 6.1 *Let $\lambda > 0$ and $v : [0, \infty) \to [0, \infty)$ an increasing, right continuous function satisfying $v(0) = 0$. The stochastic process $X_{v,\lambda} = \{X_{v,\lambda}(t) : t \geq 0\}$ is called a gamma process with shape function v and scale parameter $\lambda > 0$ if*

1. *$X_{v,\lambda}(0) = 0$ almost surely.*
2. *The stochastic process $X_{v,\lambda}$ has independent increments.*
3. *The random variable $X_{v,\lambda}(s) - X_{v,\lambda}(t)$, $s > t$ has a gamma distribution with shape parameter $v(s) - v(t)$ and scale parameter $\lambda > 0$.*

A gamma process is called stationary if the shape function v is linear. Otherwise it is called non-stationary. A stationary gamma process with shape function $v(t) = t$ and scale parameter 1 will be called standard and for notational convenience such a process is denoted by $X = \{X(t) : t \geq 0\}$. In Protter (1992) it is shown that there exists a unique modification of the standard gamma process which is càdlàg (right continuous sample paths having left-hand limits). We will always use this modification. Clearly the expectation of the random variable $X(t)$ is equal to

$$\mathbb{E}(X(t)) = \lambda^{-1}v(t),$$

while its variance is given by

$$\mathrm{Var}(X(t)) = \lambda^{-2}v(t).$$

To start with our model we assume that for a given strictly increasing continuous shape function v satisfying $v(0) = 0$ and $v(\infty) = \infty$ and a given scale parameter $\lambda > 0$ the (càdlàg) deterioration process of a new coating is given by a non-stationary gamma process $X_{v,\lambda}$. After having defined the gamma process we can now formally state for every $t > 0$ that

$$X_{v,\lambda}(t) := \text{size of the area affected by corrosion at time } t \text{ before first maintenance action.} \tag{6.1}$$

6.2.2 Effect of maintenance on deterioration

Having introduced model (6.1), we next give a description of the interaction between deterioration and (imperfect) maintenance. A maintenance action a is performed as soon as the size of the affected area exceeds a given level ρ. For simplicity it is assumed that any maintenance action takes a negligible amount of time and that such an action is chosen from a finite set A of possible actions. In our specific example we have $A = \{\text{spot repair, repainting, replacement}\}$. Introducing for every $r > 0$ the hitting time

$$T_{v,\lambda}(r) := \inf\{t \geq 0 : X_{v,\lambda}(t) > r\}$$

it follows that the random time L of the first maintenance action is given by $T_{v,\lambda}(\rho)$. Since it can be shown (see relation 5.2) that

$$T_{v,\lambda}(r) \stackrel{d}{=} v^{\leftarrow}(T(\lambda r))$$

with $T(\lambda r)$ denoting the hitting time to level λr of the standard gamma process X and $v^{\leftarrow}$ the inverse function of v, we obtain that

$$L \stackrel{d}{=} v^{\leftarrow}(T(\lambda\rho)). \tag{6.2}$$

By relation (6.2) this yields

$$\mathbb{P}\{L \leq t\} = \mathbb{P}\{T(\lambda\rho) \leq v(t)\} = \mathbb{P}\{X(v(t)) > \lambda\rho\}. \tag{6.3}$$

Now consider a maintenance policy $\Pi = (a_i)_{i\in\mathbb{N}}$, where $a_i \in A$ denotes the maintenance action chosen after level ρ is exceeded for the ith time. If the selected maintenance action is replacement, then the affected area has size zero again, while for actions belonging to the set $A_0 = \{\text{spot repair, repainting}\}$ the effect of the maintenance action on the size of the affected area is not known beforehand. To model the effect of these imperfect maintenance actions we first observe by the jump discontinuities of the sample paths of a gamma process, that the overshoot

$$W_{v,\lambda}(r) := X_{v,\lambda}(T_{v,\lambda}(r)) - r$$

beyond any level $r > 0$ is positive almost surely. Hence the size of the affected area at the first maintenance moment $T_{v,\lambda}(\rho)$ just before the first maintenance action a_1 is given by $\rho + W_{v,\lambda}(\rho)$. This is graphically shown in Figure 6.1. It is easy to see (cf. relation 5.6) that

$$W_{v,\lambda}(\rho) \stackrel{d}{=} \lambda^{-1} W(\lambda\rho) \tag{6.4}$$

with $W(r)$ denoting the overshoot of a standard gamma process at level r and this implies that the random size of the affected area at moment $T_{v,\lambda}(\rho)$ is distributed as $\rho + \lambda^{-1}W(\lambda\rho)$. In practice ρ is always much larger than the expected overshoot $\lambda^{-1}\mathbb{E}(W(\lambda\rho))$ (see Appendix 6.A). Hence to avoid complicated mathematical technicalities due to the assumption of a non-stationary gamma deterioration process and its associated discontinuous sample paths it seems realistic from a practical point of view to assume that the effect of any imperfect maintenance action will certainly annihilate the overshoot. Therefore in modelling the effect of an imperfect maintenance action on the size of the area affected by corrosion we will disregard the overshoot. As a matter of fact, in most studies the overshoot of the gamma process is not mentioned at all.

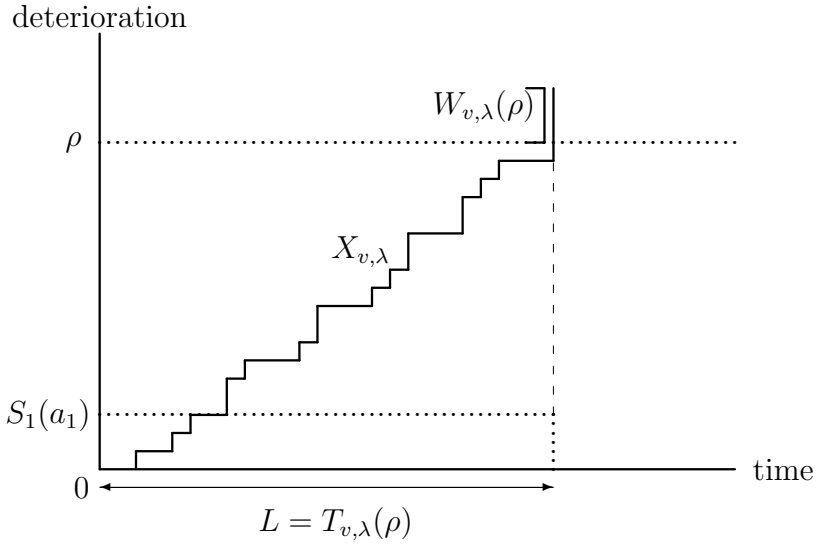

Figure 6.1: Deterioration path before the first imperfect maintenance action and the immediate effect of this action on deterioration.

The effect of the first (imperfect) maintenance action a_1 is modelled by a nonnegative random variable $S_1(a_1)$ bounded by ρ, denoting the size of the affected area just after performing the first maintenance action a_1 (see Figure 6.1). More generally, to model the impact of a first maintenance action, let us introduce the random vector $S_1 = (S_1(a))_{a \in A}$ consisting of correlated components. Within this vector the random variable $S_1(a)$ represents the size of the affected area just after performing maintenance action a. For a

representing replacement we obtain $S_1(a) = 0$ almost surely, while for a belonging to A_0 we may also allow the random variable $S_1(a)$ to take the value 0 with a small positive probability. This means that with a (possibly small) probability the effect of the imperfect maintenance actions spot repair or repainting may be the same as the effect of replacement. Since maintenance is always performed at intervention level ρ it follows that the second maintenance action a_2 in policy $\Pi = (a_i)_{i\in\mathbb{N}}$, is executed at the random time that the deterioration process occurring after the first maintenance action exceeds the (random) level $\rho - S_1(a_1)$. It is assumed that this deterioration process, independent of the previous deterioration gamma process and the (random) improvement of action a_1, is again a gamma process with continuous strictly increasing shape function v_{a_1}, satisfying $v_{a_1}(0) = 0$ and $v_{a_1}(\infty) = \infty$, and scale parameter $\lambda_{a_1} > 0$. Observe the parameters of this gamma process may depend on the first maintenance action a_1 and on the first maintenance moment.[1] Hence the second maintenance action occurs at the random time $L + L_1(a_1)$ with

$$L_1(a_1) = T_{v_{a_1},\lambda_{a_1}}(\rho - S_1(a_1)) \stackrel{d}{=} v_{a_1}^{\leftarrow}\big(T^{(1)}(\lambda_{a_1}(\rho - S_1(a_1)))\big). \tag{6.5}$$

and $T^{(1)}(r)$ the hitting time to level $r > 0$ of an independent copy $X^{(1)}$ of the standard gamma process X. This shows that the random times L and $L_1(a_1)$ are independent. Continuing in this way and disregarding the overshoot[2] of the independent copies $X^{(i)}$, $i \in \mathbb{N}$, of a standard gamma process the nth maintenance moment associated with policy $\Pi = (a_i)_{i\in\mathbb{N}}$ is distributed as the random variable

$$L + \sum_{i=1}^{n-1} L_i(a_i) \tag{6.6}$$

with L and $L_i(a_i)$, $1 \le i \le n-1$, independent nonnegative random variables and

$$L_i(a_i) \stackrel{d}{=} v_{a_i}^{\leftarrow}\big(T^{(i)}(\lambda_{a_i}(\rho - S_i(a_i)))\big). \tag{6.7}$$

Clearly in (6.7) the random variable $T^{(i)}(\cdot)$ denotes the hitting time of the standard gamma process copy $X^{(i)}$. Again, it is always assumed (this is necessary for the dynamic programming formulation to be discussed in the next section) that the parameters of the gamma deterioration process occurring between the ith and $(i+1)$th maintenance action only depend on the ith used maintenance action and the ith maintenance time. In Figure 6.2 a realization of the second maintenance moment is given. It is also assumed that for every action a the random variables $S_i(a)$, $i \in \mathbb{N}$, denoting the size of the affected

[1] This is suppressed in the notation.

[2] It is assume that in practice the improvement of the condition is always sufficiently large even for spot repair. This implies that after each imperfect maintenance action the random amount of corroded area to be counterbalanced is much larger compared to the random overshoot.

area after performing action a at the ith maintenance opportunity, are independent and identically distributed. This yields together with the assumption that for each $i \in \mathbb{N}$ the random variable $S_i(a)$ is also independent of $X^{(i)}$ (and hence $T^{(i)}(\cdot)$), that by relation (6.7) the random variables $L_i(a)$, $i \in \mathbb{N}$, are independent and identically distributed. In particular, this implies

$$L_i(a) \stackrel{d}{=} v_a^{\leftarrow}(T(\lambda_a R_a)) \tag{6.8}$$

with the random variable $R_a := \rho - S_1(a)$ concentrated on $[0, \rho]$ independent of the standard gamma process X and its associated hitting time $T(\cdot)$. In the remainder of this chapter the CDF of the random variables $L_i(a)$, $i \in \mathbb{N}$, is denoted by F_a and by relation (6.8) and v_a strictly increasing and continuous satisfying $v_a(0) = 0$ and $v_a(\infty) = \infty$ we obtain

$$F_a(t) = \mathbb{P}\{T(\lambda_a R_a) \leq v_a(t)\} = \mathbb{P}\{X(v_a(t)) > \lambda_a R_a\}. \tag{6.9}$$

Since v_a is continuous, $v_a(0) = 0$ and R_a a positive random variable independent of X with CDF G_{R_a} satisfying $G_{R_a}(0) = 0$ we obtain by the definition of a gamma process that F_a is continuous and satisfies $F_a(0) = 0$. Under some additional condition on the CDF G_{R_a} one can actually show that the CDF F_a is Lipschitz continuous. This property is needed to bound the discretization error of the continuous-time dynamic programming equation to be discussed in Section 6.3. For completeness we first list the definition of a Lipschitz continuous function.

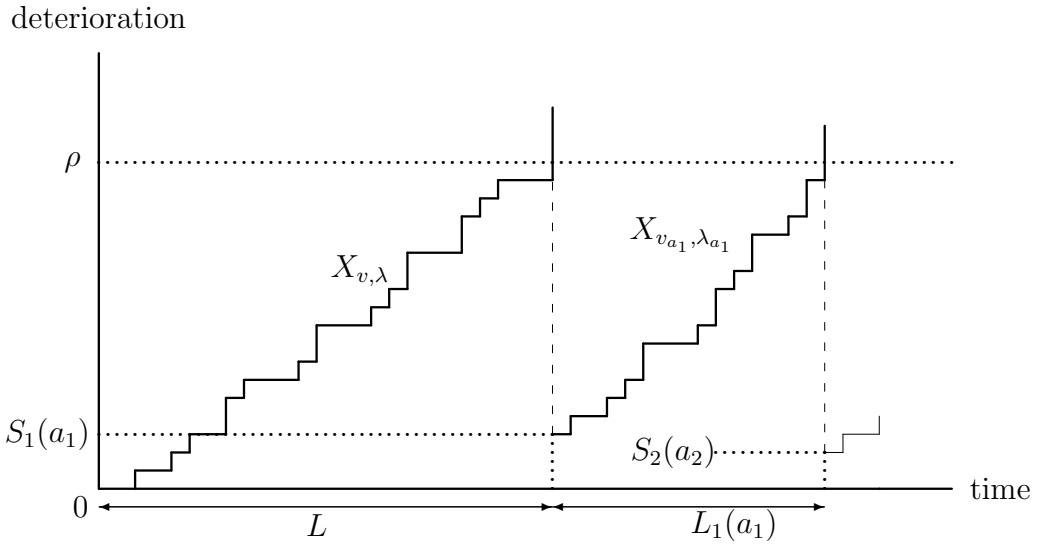

Figure 6.2: Deterioration paths before and after the first imperfect maintenance action as well as the effects of the first two maintenance actions on deterioration.

Definition 6.2 *A function $u : [0,\infty) \to \mathbb{R}$ is called Lipschitz continuous on the set $B \subseteq [0,\infty)$ with Lipschitz constant C if*

$$|u(t+s) - u(t)| \leq Cs$$

for every $t, t+s \in B$ and $s > 0$.

One can now show the following result.

Lemma 6.3 *If the CDF G_{R_a} for some $a \in A$ is Lipschitz continuous on $[0,\rho]$ with Lipschitz constant C, then*

$$0 \leq F_a(t+s) - F_a(t) \leq C\lambda_a^{-1}(v_a(t+s) - v_a(t))$$

for every $t, s > 0$.

Proof. Since the CDF G_{R_a} is Lipschitz continuous on $[0,\rho]$ and G_{R_a} is the CDF of a random variable R_a satisfying $G_{R_a}(\rho) = 1$ the CDF G_{R_a} is continuous on $[0,\infty)$. This shows $\mathbb{P}\{R_a < x\} = \mathbb{P}\{R_a \leq x\}$ for every $x > 0$ and by conditioning on $X(v_a(t)) \sim F_{v_a(t)}$ we obtain from relation (6.9) that

$$F_a(t) = \int_0^\infty \mathbb{P}\{R_a \leq x\lambda_a^{-1}\} dF_{v_a(t)}(x) = \mathbb{E}\big(G_{R_a}(\lambda_a^{-1} X(v_a(t)))\big).$$

This implies by the Lipschitz continuity of the CDF G_{R_a} and X an increasing process that

$$\begin{aligned} 0 &\leq F_a(t+s) - F_a(t) \\ &= \mathbb{E}\big(G_{R_a}(\lambda_a^{-1} X(v_a(t+s)))\big) - \mathbb{E}\big(G_{R_a}(\lambda_a^{-1} X(v_a(t)))\big) \\ &\leq C\lambda_a^{-1}\mathbb{E}\big(X(v_a(t+s)) - X(v_a(t))\big) \\ &= C\lambda_a^{-1}(v_a(t+s) - v_a(t)) \end{aligned}$$

and hence the desired inequality is verified. □

For degenerate random variables one can also show the following result.

Lemma 6.4 *It follows for $R_a = \rho$ almost surely that there exists some constant $C > 0$ satisfying*

$$F_a(t+s) - F_a(t) \leq C(v_a(t+s) - v_a(t))$$

for every $0 < t < t+s \leq T$.

Proof. If we consider a gamma process with shape function $v_a(t) = t$ for every $t > 0$ and scale parameter λ_a we obtain by relation (6.9) that

$$F_a(t) = \mathbb{P}\{X(t) \geq \lambda_a \rho\} = \frac{1}{\Gamma(t)} \int_\rho^\infty x^{t-1} \exp(-x) dx.$$

It is well known that the function $t \mapsto \Gamma(t)$ is infinitely differentiable on $(0, \infty)$ and $t\Gamma(t) = \Gamma(t+1)$ for every $t > 0$ (Rudin, 1976; Whittaker and Watson, 1958). This implies

$$t\frac{d\Gamma}{dt}(t) + \Gamma(t) = \frac{d\Gamma}{dt}(t+1)$$

and so

$$\frac{\frac{d\Gamma}{dt}(t)}{\Gamma(t)^2} = \frac{\frac{d\Gamma}{dt}(t+1)}{t\Gamma(t)^2} - \frac{1}{t\Gamma(t)} = \frac{\frac{d\Gamma}{dt}(t+1)}{\Gamma(t+1)\Gamma(t)} - \frac{1}{\Gamma(t+1)}$$

for every $t > 0$. Hence we obtain using $\lim_{t \downarrow 0} \Gamma(t) = \infty$ that

$$\lim_{t \downarrow 0} \frac{\frac{d\Gamma}{dt}(t)}{\Gamma(t)^2} = -\frac{1}{\Gamma(1)} = -1. \tag{6.10}$$

Using relation (6.10) and $t \mapsto \Gamma(t)^{-1}$ is continuously differentiable on $(0, \infty)$ this yields

$$\sup\{\frac{\frac{d\Gamma}{dt}(t)}{\Gamma(t)^2} : 0 < t \leq T\} < \infty$$

and so the function $t \mapsto \Gamma(t)^{-1}$ has a uniformly bounded derivative on $(0, T]$. A similar observation also applies to the function $t \mapsto \int_\rho^\infty x^{t-1} \exp(-x)dx$ and both functions $t \mapsto \Gamma(t)^{-1}$ and $t \mapsto \int_\rho^\infty x^{t-1} \exp(-x)dx$ are therefore Lipschitz continuous on $(0, T]$. Hence the product of these functions is also Lipschitz continuous and we have shown that there exists some $C > 0$ satisfying

$$F_a(t+s) - F_a(t) \leq Cs. \tag{6.11}$$

For general shape functions $v_a(t)$ it follows again by relation (6.9) and the result for a standard gamma process given in relation (6.11) that

$$F_a(t+s) - F_a(t) \leq C(v_a(t+s) - v_a(t))$$

and the result is verified. □

Finally we show the following result.

Lemma 6.5 *If the CDF G_{R_a} is Lipschitz continuous on $[0, \rho)$ and the CDF G_{R_a} has a jump discontinuity of size $0 < \alpha < 1$ at ρ, then there exists some $C > 0$ satisfying*

$$F_a(t+s) - F_a(t) \leq C(v_a(t+s) - v_a(t)).$$

Proof. By relation (6.9) we obtain

$$F_a(t) = \mathbb{P}\{X(v_a(t)) > \lambda_a R_a, R_a < \rho\} + \alpha \mathbb{P}\{X(v_a(t)) > \lambda_a \rho\}. \tag{6.12}$$

Observe now that by the continuity of G_{R_a} on $[0, \rho)$ and conditioning on $X(v_a(t))$ that

$$\begin{aligned} \mathbb{P}\{X(v_a(t)) > \lambda_a R_a, R_a < \rho\} &= \int_0^\infty \mathbb{P}\{\lambda_a R_a \leq \min\{x, \lambda_a \rho\}\} dF_{v_a}(x) \\ &= \int_0^\infty G_{R_a}(\min\{\lambda_a^{-1} x, \rho\}) dF_{v_a}(x) \\ &= \mathbb{E}\big(G_{R_a}(\min\{\lambda_a^{-1} X(v_a(t)), \rho\})\big). \end{aligned} \tag{6.13}$$

Since by our assumption the function $x \mapsto G_{R_a}\big(\min\{\lambda_a^{-1} x, \rho\}\big)$ is Lipschitz continuous on $[0, \infty)$ the desired result follows by applying Lemma 6.4 to (the second term of) relation (6.12) and Lemma 6.3 to relation (6.13). □

Observe the conditions in the above lemmas on the CDF G_{R_a} cover the practical case that with probability smaller than or equal to 1 a maintenance action a yields that the corroded area has size 0 again, while at the same time the CDF of the random reduction below level ρ should be Lipschitz continuous. This last condition is satisfied by a lot of conditional distribution functions truncated at level ρ. If additionally the shape function v_a is Lipschitz continuous on $[0, T]$ an immediate consequence of the above results is that the function F_a is also Lipschitz continuous on $[0, T]$. As already observed this property plays an important role in showing that the error due to discretizing the continuous-time model to be presented in the next section remains bounded. Hence with a known (small) error we may solve the discretized version of this model in the computational section.

6.3 Maintenance optimization

In this section we introduce a finite horizon optimization model for the maintenance of coating systems protecting steel structures. In Section 6.3.1 the model is formulated as a continuous-time stochastic dynamic programming problem. The corresponding renewal-type optimality equation can only be solved by discretizing time. To this end we propose a simple numerical procedure in Section 6.3.2 and we investigate in detail the accuracy of this procedure.

6.3.1 Continuous-time model

Let the length of the total planning horizon be denoted by T. Observe T represents in practice the finite usage time of the steel structure. The maintenance optimization problem is about which action to select when the deterioration exceeds the fixed intervention level ρ. The aim is to minimize the expected maintenance costs over the (finite) planning horizon with respect to the policy $\Pi = \{a_i\}_{i \in \mathbb{N}}$. The cost of a given maintenance action a is denoted by $c(a)$ and it does not depend on (the parameters of) the deterioration process.

To formulate the associated Bellman (optimality) equation of the above dynamic programming problem, let $1_{\mathcal{A}} : [0,T] \to \mathbb{R}$ be the indicator function of the set $\mathcal{A} \subseteq \mathbb{R}$ given by

$$1_{\mathcal{A}}(t) = \begin{cases} 1 & \text{if } t \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}.$$

Moreover, denote by $q(t)$, $0 \leq t \leq T$, the (conditional) minimal expected maintenance costs from time $T-t$ up to time T given that at time $T-t$ a crossing occurs and hence maintenance needs to be done. If this crossing happens for the i^*th time and we select at time $T-t$ action $a \in A$, then due to the randomness of the next maintenance moment determined by our chosen action a the conditional minimal (random) maintenance costs from time $T-t$ up to time T is given by

$$c(a)1_{(0,T]}(t) + q\big(\max\{t - L_{i^*}(a), 0\}\big) \tag{6.14}$$

with $q(0) = 0$. To justify relation (6.14) observe for $L_{i^*}(a) > t$ that the present maintenance action will be the last one and so the total costs from $T-t$ up to T equal $c(a)1_{(0,T]}(t)$. For $t = 0$ we are at the end of the planning horizon and so we do not need to perform a maintenance action anymore. This leads to zero maintenance cost explaining the indicator function in the above expression. If $L_{i^*}(a) \leq t$ the next maintenance action occurs at time $T - (t - L_{i^*}(a))$ and so in this case the conditional minimal (random) maintenance costs from $T-t$ up to T equal

$$c(a)1_{(0,T]}(t) + q(t - L_{i^*}(a)).$$

The last term occurs since we need to select from time $T - (t - L_{i^*}(a))$ up to T in an optimal way the (possible) remaining actions. By construction the random variables $L_i(a)$, $i \in \mathbb{N}$, are independent and identically distributed and for each i the random variable $L_i(a)$ is also independent of the maintenance costs occurring after the $(i+1)$th maintenance moment. Hence from relation (6.14) we obtain that the conditional *expected* maintenance costs from time $T-t$ up to T given action a is selected equal

$$c(a)1_{(0,T]}(t) + \mathbb{E}\big(q\big(\max\{t - L_{i^*}(a), 0\}\big)\big) = c(a)1_{(0,T]}(t) + \mathbb{E}\big(q\big(\max\{t - L_1(a), 0\}\big)\big). \tag{6.15}$$

Selecting now at moment $T-t$ the best possible maintenance action we obtain

$$q(t) = \min_{a \in A}\{c(a)1_{(0,T]}(t) + \mathbb{E}(q(\max\{t - L_1(a), 0\}))\} \tag{6.16}$$

for every $0 \leq t \leq T$. Introducing the convolution operation $*$ given by

$$(q * F_a)(t) := \int_0^t q(t-y)dF_a(y) \tag{6.17}$$

for every $t \geq 0$ relation (6.16) reduces to

$$q(t) = \min_{a \in A} \left\{ c(a) 1_{(0,T]}(t) + (q * F_a)(t) \right\} \tag{6.18}$$

for every $0 \leq t \leq T$. By relation (6.18) and its definition the optimal value function $q : [0, T] \to \mathbb{R}$ is increasing, satisfies $q(0) = 0$ and has a jump at zero. In the next subsection it is also shown that the function q is continuous on $(0, T]$. Unfortunately it is difficult to solve the above continuous-time optimality equation and so we need to discretize this equation to find the optimal maintenance policy.

If we consider within the same finite time horizon T the discounted cost objective function with discount rate $\delta > 0$, then we proceed as follows. Let $q_\delta(t)$ denote the conditional minimal expected total discounted costs from $T - t$ up to T, evaluated at $T - t$, if maintenance needs to be performed at $T - t$ and the discount factor is given by δ. By a similar argument as for the non-discounted case we obtain

$$q_\delta(t) = \min_{a \in A} \{ c(a) 1_{(0,T]}(t) + \int_0^t \exp(-\delta y) q_\delta(t - y) dF_a(y) \}. \tag{6.19}$$

Hence it follows

$$q_\delta(t) = \min_{a \in A} \{ c(a) 1_{(0,T]}(t) + (q_\delta * F_{a,\delta})(t) \} \tag{6.20}$$

with $F_{a,\delta}$ the defective cdf

$$F_{a,\delta}(t) := \int_0^t \exp(-\delta y) dF_a(y). \tag{6.21}$$

Since the structure of the optimality equation in relation (6.20) is the same as the optimality equation in relation (6.18) both relations can be solved in a similar way by discretizing. We will only discuss in detail the discretization procedure applied to relation (6.18). By the above definition it follows that the minimal expected total discounted costs $q_{\delta,0}(t)$ from $T - t$ up to T, evaluated at time 0, if maintenance needs to be performed at time $T - t$ is given by

$$q_{\delta,0}(t) = \exp(-\delta(T - t)) q_\delta(t). \tag{6.22}$$

Notice the resemblance between Equations (6.18) and (6.19) and the optimality equations in Section 5 in Newby and Dagg (2004) and in Section 5.2 in Newby and Barker (2006). The difference being that in their model the action space (inspection time) is also continuous. Under the assumption that the (discounted) cost objective function is continuous and bounded they discretize the state space and apply a policy improvement algorithm to find the optimal aperiodic inspection policy. In Section 6.3.2 we will show that the above optimality equations are (Lipschitz) continuous. Moreover, we show that it is easy to solve the discretized versions of these equations for different values of $0 < t \leq T$ without policy improvement.

Finally, if the horizon is infinite, then the minimal total expected discounted costs equal $\mathbb{E}(\exp(-\delta L))q_\delta^\infty$ with L listed in relation (6.2) and q_δ^∞ denoting the minimal total expected discounted costs starting from, and evaluated at, the first maintenance moment. It is easy to see (cf. Eq. (6.19) as $t \to \infty$) that q_δ^∞ satisfies the recurrence relation

$$q_\delta^\infty = \min_{a \in A}\{c(a) + q_\delta^\infty \mathbb{E}(\exp(-\delta L_1(a)))\} \tag{6.23}$$

with $\mathbf{L}_1(a)$ listed in relation (6.5) and this implies

$$q_\delta^\infty = \min_{a \in A}\left\{\frac{c(a)}{1 - \mathbb{E}(\exp(-\delta L_1(a)))}\right\}. \tag{6.24}$$

Hence for the infinite horizon model with discounting the optimal stationary policy is given by that action which minimizes the expression in relation (6.24). If the decision criterion is the expected average cost per unit time, then the optimal stationary policy is given by that action that minimizes $c(a)/\mathbb{E}(L_1(a))$ (the expected average cost when action a is always selected).

6.3.2 Solving a discrete version of the renewal-type optimality equation

In this section we solve a discrete version of the continuous-time optimality equation listed in relation (6.18). In particular, we focus on the Riemann lower and upper sum approximation of this renewal-type integral equation. To start with the analysis of the discretization error let $\mathcal{B}$ denote the set of bounded functions $u : [0, T] \to \mathbb{R}$ integrable with respect to F_a, $a \in A$. If

$$\|u\|_{T,\infty} := \sup\{|u(t)| : 0 \le t \le T\}$$

denotes the well-known supnorm on $\mathcal{B}$, the vector space $(\mathcal{B}, \|.\|_{T,\infty})$ is a Banach space (Kreyszig, 1978). On this Banach space we introduce the operator $P : \mathcal{B} \to \mathcal{B}$ given by

$$Pu(t) := \min_{a \in A}\left\{c(a)1_{(0,T]}(t) + \int_0^t u(t-y)dF_a(y)\right\}$$

for every $0 \le t \le T$. To discretize the above operator let $h > 0$ be chosen in such a way that $T = Nh$ for some $N \in \mathbb{N}$ and introduce $t_k := kh$, $k = 0, \ldots, N$, and

$$p_k^{(a)} := F_a(t_{k+1}) - F_a(t_k)$$

for $a \in A$ and $k = 0, \ldots, N-1$. If for $i = 0, 1$ we introduce the set $\mathcal{B}_i := \{(u(t_i), \ldots, u(t_N)) : u \in \mathcal{B}\}$ and this vector space has Chebyshev norm

$$\|u\|_{i,d} := \max\{|u(t_n)| : n = i, \ldots, N\},$$

let the operator $U_h : \mathcal{B}_1 \to \mathcal{B}_1$ be given by

$$U_h u(t_n) := \min_{a \in A} \Big\{ c(a) + \sum\nolimits_{k=0}^{n-1} u(t_{n-k}) p_k^{(a)} \Big\} \tag{6.25}$$

for $n = 1, \ldots, N$. Also introduce the operator $L_h : \mathcal{B}_0 \to \mathcal{B}_0$ given by

$$L_h u(t_n) := \min_{a \in A} \Big\{ c(a) + \sum\nolimits_{k=0}^{n-1} u(t_{n-k-1}) p_k^{(a)} \Big\} \tag{6.26}$$

for every $n = 1, \ldots, N$ and $L_h u(t_0) = 0$. Before discussing some properties of the above operators we need the following important observation. By relation (6.9) and X is a standard gamma process, it follows

$$F_a(T) = \mathbb{P}\{X(v_a(T)) > \lambda_a R_a\} < 1$$

and since A is a finite action set this implies

$$\sigma := \max\nolimits_{a \in A} \{F_a(T)\} < 1. \tag{6.27}$$

Before mentioning the next result, observe an operator K is called a contraction (with respect to a given norm $\|.\|$) if there exists some $0 < \beta < 1$ such that

$$\|Ku - Kw\| \leq \beta \|u - w\|$$

for every u, v belonging to the domain of K. The value $0 < \beta < 1$ is called the contraction number.

Lemma 6.6 *The operator $P : \mathcal{B} \to \mathcal{B}$ is a contraction (with respect to $\|.\|_{T,\infty}$) and its contraction number is given by $0 < \sigma < 1$. The same holds for the operator $L_h : \mathcal{B}_0 \to \mathcal{B}_0$ with respect to the norm $\|.\|_{0,d}$ and the operator $U_h : \mathcal{B}_1 \to \mathcal{B}_1$ with respect to the norm $\|.\|_{1,d}$ and both contractions have the same contraction number $0 < \sigma < 1$.*

Proof. We only give a proof for the operator P, since the proof for the other operators is similar. Let $u, w \in \mathcal{B}$ be given and $0 \leq t \leq T$ fixed. If $a_w \in \arg\min_{a \in A}\{c(a) + \int_0^t w(t - y) dF_a(y)\}$ it follows by the definition of P that

$$Pu(t) - Pw(t) \leq \int_0^t (u - w)(t - y) dF_{a_w}(y). \tag{6.28}$$

Also, for $a_u \in \arg\min_{a \in A}\{c(a) + \int_0^t u(t - y) dF_a(y)\}$ we obtain similarly

$$Pu(t) - Pw(t) \geq \int_0^t (u - w)(t - y) dF_{a_u}(y). \tag{6.29}$$

By relations (6.28) and (6.29) and the definition of σ given in relation (6.27) this implies

$$|Pu(t) - Pw(t)| \leq \|u - w\|_{T,\infty} \ \max\{F_{a_w}(T), F_{a_u}(T)\} \leq \sigma \|u - w\|_{T,\infty}.$$

Since the above inequality holds for every $0 \leq t \leq T$ we obtain the desired result. □

Since $(\mathcal{B}, \|.\|_{T,\infty})$ is a Banach space it follows by Lemma 6.6 and the Banach fixed point theorem (Kreyszig, 1978) that the operator P has a unique fixed point q and for every $u \in \mathcal{B}$ the sequence $P^m u$ with $P^m u := P(P^{m-1}u)$ converges in the supnorm to this fixed point q. Hence we obtain for every $u \in \mathcal{B}$ that

$$\lim_{m\uparrow\infty} \|P^m u - q\|_{T,\infty} = 0 \text{ and } q = Pq. \tag{6.30}$$

By relation (6.18) this fixed point q represents the optimal value function of our dynamic program. For the operator L_h and U_h we obtain similarly

$$\lim_{m\uparrow\infty} L_h^m u = \underline{q} \text{ and } \lim_{m\uparrow\infty} U_h^m u = \overline{q} \tag{6.31}$$

with $\underline{q}$, respectively $\overline{q}$, the unique fixed point of the operator L_h, respectively U_h. This means $\underline{q}(t_n) = L_h \underline{q}(t_n)$ for every $n = 0, \ldots, N$ and $\overline{q}(t_n) = U_h \overline{q}(t_n)$ for every $n = 1, \ldots, N$. Observe the fixed point $\underline{q}$ of the operator L_h is easy to compute by forward substitution. Also at the end of this section we show how to compute $\overline{q}$. Using relations (6.30) and (6.31) and induction on n it is easy to show that the sequences $(\underline{q}(t_n))_{n=1}^N$ and $(\overline{q}(t_n))_{n=1}^N$ are resp. lower and upper bounds for the sequence $(q(t_n))_{n=1}^N$.

Lemma 6.7 *For every $n = 1, \ldots, N$ it follows $\underline{q}(t_n) \leq q(t_n) \leq \overline{q}(t_n)$.*

Proof. Let $u \in \mathcal{B}$ be increasing. Since u is increasing we obtain for every $n = 1, \ldots, N$ that

$$\begin{aligned} \int_0^{t_n} u(t_n - y) dF_a(y) &= \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} u(t_n - y) dF_a(y) \\ &\leq \sum_{k=0}^{n-1} u(t_n - t_k) p_k^{(a)} \\ &= \sum_{k=0}^{n-1} u(t_{n-k}) p_k^{(a)}. \end{aligned}$$

This shows by the definition of P and U_h that

$$Pu(t_n) \leq U_h u(t_n)$$

for every $n = 1, \ldots, N$. Suppose now by induction that $P^m u(t_n) \leq U_h^m u(t_n)$ for some $m \in \mathbb{N}$ and $n = 1, \ldots, N$. Since it is easy to verify that $t \to P^m u(t)$ is increasing for every

increasing $u \in \mathcal{B}$ this implies

$$\begin{aligned} P^{m+1}u(t_n) &= \min_{a\in A}\left\{c(a) + \int_0^{t_n} P^m u(t_n - y)dF_a(y)\right\} \\ &= \min_{a\in A}\left\{c(a) + \sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}} P^m u(t_n - y)dF_a(y)\right\} \\ &\leq \min_{a\in A}\left\{c(a) + \sum_{k=0}^{n-1} P^m u(t_{n-k})p_k^{(a)}\right\} \\ &\leq \min_{a\in A}\left\{c(a) + \sum_{k=0}^{n-1} U_h^m u(t_{n-k})p_k^{(a)}\right\} \\ &= U_h^{m+1}u(t_n). \end{aligned}$$

Hence we have verified that $P^m u(t_n) \leq U_h^m u(t_n)$ for every $m \in \mathbb{N}$ and $n = 1, \ldots, N$. This implies by relations (6.30) and (6.31) that

$$q(t_n) = \lim_{m\uparrow\infty} P^m u(t_n) \leq \lim_{m\uparrow\infty} U_h^m u(t_n) = \overline{q}(t_n).$$

By a similar proof one can show that $q(t_n) \geq \underline{q}(t_n)$ for $n = 1, \ldots, N$ and the result is verified. □

In the next result we show that the fixed point q of the operator P is a continuous function on (0,T] and under some additional condition even Lipschitz continuous on this set. Observe Lipschitz continuity of the fixed point q on $(0, T]$ is important for the determination of an upper bound on the discretization error.

Lemma 6.8 *The fixed point q of the operator P is continuous on $(0, T]$ and has a jump discontinuity at 0. Moreover, if F_a is Lipschitz continuous on $[0, T]$ for every $a \in A$, then q is also Lipschitz continuous on $(0, T]$.*

Proof. Since the function q is increasing it is sufficient to construct an upper bound on $q(t + s) - q(t)$ with $0 < t < t + s \leq T$. To start the proof introduce for every increasing and bounded function $u : [0, \infty) \to [0, \infty)$

$$d_s(u) := \sup_{0<x\leq T-s}\{u(x + s) - u(x)\}. \tag{6.32}$$

By the definition of the operator P, $q = Pq$ and q increasing we obtain for every $0 < t \leq T - s$ fixed that there exists some $a(t) \in A$ (possibly depending on t) satisfying

$$\begin{aligned} q(t + s) - q(t) &\leq \int_0^t (q(t + s - y) - q(t - y))dF_{a(t)}(y) + \int_t^{t+s} q(t + s - y)dF_{a(t)}(y) \\ &\leq \int_0^t (q(t + s - y) - q(t - y))dF_{a(t)}(y) + q(s)\max_{a\in A}\{d_s(F_a)\}. \end{aligned} \tag{6.33}$$

Since by the observation after relation (6.9) the CDF $F_{a(t)}$ is continuous it follows that

$$\int_0^t (q(t+s-y)-q(t-y))dF_{a(t)}(y) = \lim_{p\uparrow 0}\int_0^{t-p}(q(t+s-y)-q(t-y))dF_{a(t)}(y). \quad (6.34)$$

Using $0 < t < T - s$ and relation (6.32) we obtain for every $p > 0$ that

$$\int_0^{t-p}(q(t+s-y)-q(t-y))dF_{a(t)}(y) \leq d_s(q)F_{a(t)}(T).$$

Hence by relations (6.34) and (6.27)

$$\int_0^t (q(t+s-y)-q(t-y))dF_{a(t)}(y) \leq \sigma d_s(q).$$

Using the above inequality and relation (6.33) yields

$$q(t+s)-q(t) \leq \sigma d_s(q) + q(s)\max_{a\in A}\{d_s(F_a)\}. \quad (6.35)$$

Since relation (6.35) holds for every $0 < t < t+s \leq T$, we finally obtain that

$$d_s(q) \leq \sigma d_s(q) + q(s)\max_{a\in A}\{d_s(F_a)\}.$$

By relation (6.27) we know that $\sigma < 1$ and so

$$d_s(q) \leq \frac{q(s)\max_{a\in A}\{d_s(F_a)\}}{1-\sigma}. \quad (6.36)$$

Since the CDF F_a is continuous on $[0,\infty)$ and hence uniformly continuous on $[0,T]$ (Rudin, 1976), implying

$$\lim_{s\downarrow 0} d_s(F_a) = 0$$

for every $a \in A$, the continuity of q on $(0,T)$ follows by relation (6.36) and A finite. For F_a, $a \in A$, Lipschitz continuous we obtain for every $a \in A$ that there exists some finite $C_a > 0$ satisfying

$$d_s(F_a) \leq C_a s.$$

Again by relation (6.36) and A finite the Lipschitz continuity of q on $(0,T]$ follows. □

In the next lemma we give a conservative bound on the error $\|q - \underline{q}\|_{1,d}$ if the CDF's F_a, $a \in A$, are Lipschitz continuous.

Lemma 6.9 *If the CDF F_a is Lipschitz continuous for every $a \in A$, then there exists some constant $C > 0$ independent of h such that*

$$\|q - \underline{q}\|_{1,d} \leq Ch.$$

Proof. By the triangle inequality and q, respectively $\underline{q}$, is a fixed point of the operator P, respectively L_h, we obtain

$$\begin{aligned} \|q - \underline{q}\|_{1,d} &= \|Pq - L_h\underline{q}\|_{1,d} \\ &\leq \|Pq - L_hq\|_{1,d} + \|L_hq - L_h\underline{q}\|_{1,d}. \end{aligned} \tag{6.37}$$

Since F_a is Lipschitz continuous for every $a \in A$ we obtain by Lemma 6.8 that the fixed point q is Lipschitz continuous and increasing on $(0, T]$ (remember $q(0) = 0$) and so there exists some finite positive constant C_q satisfying

$$|Pq(t_n) - L_hq(t_n)| \leq C_qh + q(h) \max_{a\in A}\{p_{n-1}^{(a)}\} \tag{6.38}$$

for every $n = 1, \ldots, N$. Introducing for a CDF F the value

$$d_h(F) := \sup_{k=0,\ldots,N-1}\{F(t_{k+1}) - F(t_k)\} \tag{6.39}$$

it follows by relation (6.38) that

$$\|Pq - L_hq\|_{1,d} \leq C_qh + q(h) \max_{a\in A}\{d_h(F_a)\}. \tag{6.40}$$

Since L_h is a contraction with contraction number $0 < \sigma < 1$ and $q(0) = \underline{q}(0) = 0$ this implies by relations (6.37) and (6.40) that

$$\|q - \underline{q}\|_{1,d} \leq C_qh + q(h) \max_{a\in A}\{d_h(F_a)\} + \sigma\|q - \underline{q}\|_{1,d}$$

and so

$$\|q - \underline{q}\|_{1,d} \leq \frac{C_qh + q(h)\max_{a\in A}\{d_h(F_a)\}}{1 - \sigma}. \tag{6.41}$$

By the Lipschitz continuity of F_a with Lipschitz constant C_a we obtain $d_h(F_a) \leq C_ah$. Also by the Lipschitz continuity of q on $(0, T]$ it follows that

$$q(h) = q(h) - \lim_{t\downarrow 0} q(t) + \min_{a\in A}\{c(a)\} \leq C_qh + \min_{a\in A}\{c(a)\} \tag{6.42}$$

and this shows in combination with relation (6.41) and A a finite set the desired result.

□

It is also possible without any conditions on the CDF F_a to obtain an estimate of the supnorm error. Introducing

$$\nu_h := \max_{1\leq k\leq N-1}\{\underline{q}(t_{k+1}) - \underline{q}(t_k)\}$$

one can show the following result.

Lemma 6.10 *It follows that*

$$\|q - \underline{q}\|_{1,d} \leq \|\overline{q} - \underline{q}\|_{1,d} \leq \frac{\nu_h + \underline{q}(t_1) \max_{a \in A} \{d_h(F_a)\}}{1 - \sigma}$$

with $d_h(F_a)$ defined in relation (6.39).

Proof. For every $n = 1, \ldots, N$ that there exists some $a_n \in A$ such that

$$U_h \underline{q}(t_n) - L_h \underline{q}(t_n) \leq \sum_{k=0}^{n-1} \underline{q}(t_{n-k}) p_k^{(a_n)} - \sum_{k=0}^{n-1} \underline{q}(t_{n-k-1}) p_k^{(a_n)}$$

Using $\underline{q}(0) = 0$ the last term can be rewritten as

$$\sum_{k=0}^{n-1} \underline{q}(t_{n-k}) p_k^{(a_n)} - \sum_{k=0}^{n-1} \underline{q}(t_{n-k-1}) p_k^{(a_n)} = \sum_{k=0}^{n-2} (\underline{q}(t_{n-k}) - \underline{q}(t_{n-k-1})) p_k^{(a_n)} + \underline{q}(t_1) p_{n-1}^{(a_n)}.$$

Since $\underline{v}$ is increasing (check this by using $\lim_{m \uparrow \infty} L_h^m u(t_n) = \underline{v}(t_n)$ and u increasing implies $L_h u$ increasing) we therefore obtain

$$U_h \underline{q}(t_n) - L_h \underline{q}(t_n) \leq \nu_h + \underline{q}(t_1) \max_{a \in A} \{p_{n-1}^{(a)}\}.$$

This implies

$$\|U_h \underline{q} - L_h \underline{q}\|_{1,d} \leq \nu_h + \underline{q}(t_1) \max_{a \in A} \{d_h(F_a)\}. \tag{6.43}$$

Also, using U_h is a contraction with contraction number σ, it follows

$$\begin{aligned} \|U_h \overline{q} - L_h \underline{q}\|_{1,d} &= \|U_h \overline{q} - U_h \underline{q} + U_h \underline{q} - L_h \underline{q}\|_{1,d} \\ &\leq \|U_h \overline{q} - U_h \underline{q}\|_{1,d} + \|U_h \underline{q} - L_h \underline{q}\|_{1,d} \\ &\leq \sigma \|\overline{q} - \underline{q}\|_{1,d} + \|U_h \underline{q} - L_h \underline{q}\|_{1,d}. \end{aligned} \tag{6.44}$$

Combining relations (6.43) and (6.44) finally yields

$$\begin{aligned} \|\overline{q} - \underline{q}\|_{1,d} &= \|U_h \overline{q} - L_h \underline{q}\|_{1,d} \\ &\leq \sigma \|\overline{q} - \underline{q}\|_{1,d} + \nu_h + \underline{q}(t_1) \max_{a \in A} \{d_h(F_a)\} \end{aligned}$$

and hence

$$\|\overline{q} - \underline{q}\|_{1,d} \leq \frac{\nu_h + \underline{q}(t_1) \max_{a \in A} \{d_h(F_a)\}}{1 - \sigma}. \tag{6.45}$$

By Lemma 6.7 it also follows that

$$\|q - \underline{q}\|_{1,d} \leq \|\overline{q} - \underline{q}\|_{1,d}$$

and hence by relation (6.45) we have shown the result. □

As mentioned earlier, it is easy to compute the fixed point $\underline{q}$ of the operator L_h by forward substitution. Since the right-hand side of equation (6.25) involves a function evaluation in t_n, it seems to be more complicated to compute the fixed point $\overline{q}$ of U_h. However, it turns out that there exists an easy analytic expression for $\overline{q}(t_n)$, $n = 1, 2, \ldots, N$. To this end we require the following lemma.

Lemma 6.11 *For A a finite set, $f : A \to [0,\infty)$ and $g : A \to [0,1)$ the optimality equation*

$$w = \min_{a\in A} \{f(a) + g(a)w\}. \tag{6.46}$$

has a unique solution $w \geq 0$ given by

$$w = \min_{a\in A} \left\{ \frac{f(a)}{1-g(a)} \right\}. \tag{6.47}$$

Proof. Since $w \mapsto \min_{a\in A}\{f(a) + g(a)w\}$ is a contraction, the above optimality equation has a unique solution. Moreover, by the definition of w we obtain $w \leq f(a) + g(a)w$ for every $a \in A$ and so

$$w \leq \min_{a\in A} \left\{ \frac{f(a)}{1-g(a)} \right\}.$$

Again by its definition there exists some $a \in A$ such that $w = f(a) + g(a)w$ and the result is proved. □

Lemma 6.12 *The fixed point $\overline{q}$ of the operator U_h listed in relation* (6.25) *is given by*

$$\overline{q}(t_1) = \min_{a\in A} \left\{ \frac{c(a)}{1-p_0^{(a)}} \right\}$$

and for $n = 2, \ldots, N$

$$\overline{q}(t_n) = \min_{a\in A} \left\{ \frac{c(a) + \sum_{k=1}^{n-1} \overline{q}(t_{n-k}) p_k^{(a)}}{1-p_0^{(a)}} \right\}.$$

Proof. Since $0 \leq p_0^{(a)} < 1$ for every a and A is a finite set the expression for $\overline{q}(t_1)$ is a direct consequence of Lemma 6.11. Also by the definition $\overline{q}(t_n)$ we obtain

$$\overline{q}(t_n) = \min_{a\in A} \left\{ c(a) + \sum\nolimits_{k=1}^{n-1} \left(\overline{q}(t_{n-k}) p_k^{(a)} \right) + p_0^{(a)} \overline{q}(t_n) \right\}$$

for every $n = 2, \ldots, N$. Taking $f(a) = c(a) + \sum_{k=1}^{n-1} \overline{q}(t_{n-k}) p_k^{(a)}$ and $g(a) = p_0^{(a)}$, the second formula follows again from Lemma 6.11. □

In Section 6.5 we will compute both $\underline{q}$ and $\overline{q}$ and this yields by Lemma 6.7 an upper bound on the 'empirical' accuracy of the discretization procedure(s). Also in Appendix 6.B we list a simple algorithm for computing the fixed points $\underline{q}$ and $\overline{q}$.

6.4 Computing the CDF of the time between two maintenance actions

To solve the optimization problems introduced in Section 6.3 we need a fast method to compute the CDF F_a listed in relation (6.9) of the time between two maintenance actions. We have seen in Chapter 5 that this CDF only has a nice analytical expression in some special cases. In general, evaluating this CDF numerically, e.g. via its two-dimensional integral representation, is time-consuming. However, as we will see it is easy to approximate this CDF.

Note that F_a relates to the CDF of the first time a standard gamma process exceeds a random threshold. For notational convenience we suppress the subscript a in this section and consider the CDF H_R given by $H_R(t) := \mathbb{P}\{T(R) \leq t\}$ for $t \geq 0$, where $T(R)$ is the first time a standard gamma process exceeds some nonnegative random variable R. In particular, taking $R = \lambda_a R_a$ and replacing t by $v_a(t)$ yields relation (6.9). In the remainder of this section we briefly come back to the computation of the CDF H_R for nonnegative random variables R having a general distribution. In Section 6.4.1 we give expressions for this CDF and in Section 6.4.2 we demonstrate how the desired hitting time distribution can be approximated.

6.4.1 General expressions

Since the nonnegative, non-defective, random variable R is by definition independent of the gamma process and its CDF G_R satisfies $G_R(0) = 0$ it follows by conditioning on the random variable R that

$$H_R(t) = \mathbb{P}\{T(R) \leq t\} = \mathbb{P}\{X(t) > R\} = \int_0^\infty \mathbb{P}\{X(t) > r\} dG_R(r) \tag{6.48}$$

for every $t \geq 0$. Moreover, by conditioning on the random variable $X(t)$ we obtain for G_R a continuous CDF the equivalent representation

$$H_R(t) = \mathbb{E}\big(G_R(X(t))\big). \tag{6.49}$$

In general, the above expressions have to be computed via numerical integration. On the other hand, if the random variable R has a degenerate, uniform or gamma-type distribution, then H_R has a 'nice' analytical expression (see Section 5.4).

6.4.2 A simple approximation

In Chapter 5 it was shown empirically and theoretically that a linear interpolation of H_R at the integer points approximates the true CDF quite well in the supnorm. Recall that

this approximation is given by

$$H_R(t) \approx (t - \lfloor t \rfloor) H_R(\lfloor t \rfloor + 1) + (1 - (t - \lfloor t \rfloor)) H_R(\lfloor t \rfloor), \ t \geq 0. \tag{6.50}$$

To evaluate the continuous CDF F on its integer points one uses that

$$\begin{aligned} H_R(n+1) - H_R(n) &= \mathbb{P}\{n < T(R) \leq n+1\} \\ &= \mathbb{P}\{\lfloor T(R) \rfloor = n\} \\ &= \tfrac{1}{n!}\mathbb{E}\big(R^n \exp(-R)\big) \end{aligned} \tag{6.51}$$

for every $n \in \mathbb{N} \cup \{0\}$. Alternatively, if

$$\widehat{G}_R(\tau) := \mathbb{E}(\exp(-\tau R))$$

is the probability Laplace Stieltjes transform (pLSt) of the CDF G_R, then relation (6.51) is the same as

$$H_R(n+1) - H_R(n) = \frac{(-1)^n}{n!} \widehat{G}_R^{(n)}(1) \tag{6.52}$$

for every $n \in \mathbb{N} \cup \{0\}$ with $\widehat{G}_R^{(n)}$, $n \in \mathbb{N}$, denoting the nth derivative of $\widehat{G}_R$ and $\widehat{G}_R^{(0)} := \widehat{G}_R$. If the derivatives of $\widehat{G}_R$ are elementary functions we can directly apply relation (6.52). Examples are given by the class of infinitely divisible distributions, including the gamma distribution, (power transformations of) the uniform distribution and the class of concave distributions (see Section 5.3).

Remark 6.13 *Observe the above piecewise linear approximation is derived for a standard gamma process and by construction it is Lipschitz continuous. By relation (6.3) it can also be used for a non-stationary gamma process. Approximating now the CDF of the time between two maintenance actions F_a by F_a^{prox} yields a different continuous-time dynamic programming equation, given by*

$$q_{prox}(t) = \min_{a \in A} \{c(a) 1_{(0,T]}(t) + \int_0^t q_{prox}(t-y) dF_a^{prox}(y)\}. \tag{6.53}$$

If the shape functions v_a are Lipschitz continuous on [0,T], it follows by the remark after Lemma 6.3 that the approximation F_a^{prox} is clearly Lipschitz continuous and so we may conclude from Lemma 6.8 that q_{prox} being the solution of the above approximate Bellman equation is Lipschitz continuous on $(0,T]$ and satisfies $q_{prox}(0) = 0$. Also, since F_a^{prox} is close in the supnorm to F_a, the same holds for the fixed point q_{prox} of relation (6.53) and the fixed point q of relations (6.18) and (6.19)*. By the previous results one may therefore conclude that the discretization of the above approximative Bellman equation yields accurate results.*

6.5 Numerical example

Let us illustrate the model and the methods discussed in the previous sections with some numerical examples. All computations are done in MATLAB 7.2 on a Pentium III–2 GHz personal computer. Consider a planning horizon of $T = 50$ time units. We are interested in the optimal maintenance decisions during this horizon. Let the initial gamma deterioration process be given by $X_{v,\lambda}$ with $v(t) = 0.25t^2$ and $\lambda = 1$. The intervention level is given by $\rho = 25$. Let $A = \{a_s, a_r, a_f\}$ be the set of maintenance actions representing spot repair, repainting and full replacement. Suppose the size of the affected area just after each of these three maintenance actions is given by $S(a_s) \sim \text{unif}(15, 20)$, $S(a_r) \sim \text{unif}(10, 15)$ and $S(a_f) = 0$ almost surely, respectively. It follows that the reduction R_a is uniformly distributed for $a \in \{a_s, a_r\}$ and R_{a_f} has a degenerate distribution. In this particular case F_a, $a \in A$, can be computed via an expression involving incomplete gamma functions. However, from a computational point of view the interpolation approximation introduced in the previous section is preferred as $H_R(n)$, $n \in \mathbb{N}$, satisfies a simple recurrence equation when R is uniformly distributed or degenerate (see Section 5.3) approximation is convenient in the discrete version of the optimality equations since we actually discretize the CDF. The graphs of the approximations of F_a, $a \in A$, are shown in Figure 6.3. The computation of these approximative CDFs in 5000 points takes about 0.8 seconds (vs 4.7 seconds for the exact CDFs) for $a \in \{a_s, a_r\}$ and a few milliseconds for $a = a_f$ (vs 1.5 seconds for the exact CDF).

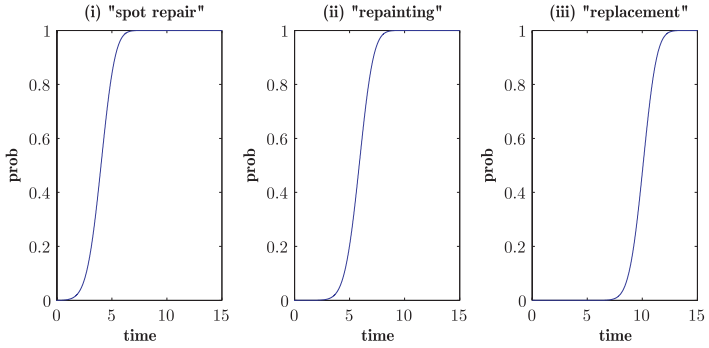

Figure 6.3: Probability distributions of the time between two maintenance actions: (i) F_{a_s}, (ii) F_{a_r}, and (iii) F_{a_f}.

The costs associated with the maintenance actions are $c(a_s) = 2$, $c(a_r) = 3$ and $c(a_f) = 5$, respectively. Finally, let the gamma processes describing the size of the affected area after maintenance action a have shape function $v_a = v$, for every action a, and scale parameter $\lambda_{a_s} = \lambda/2$, $\lambda_{a_r} = 2\lambda/3$ and $\lambda_{a_f} = \lambda$, respectively. By doing so de-

terioration is accelerated by a factor 2 (3/2) after spot repair (repainting), while keeping the variability of the deterioration process the same. Moreover, the ratio of the cost of a maintenance action and the expected time until the next crossing of level ρ due to an action is approximately 1/2 for all maintenance actions. This ratio is smallest for the action 'replacement', which minimizes the expected average cost per unit time.

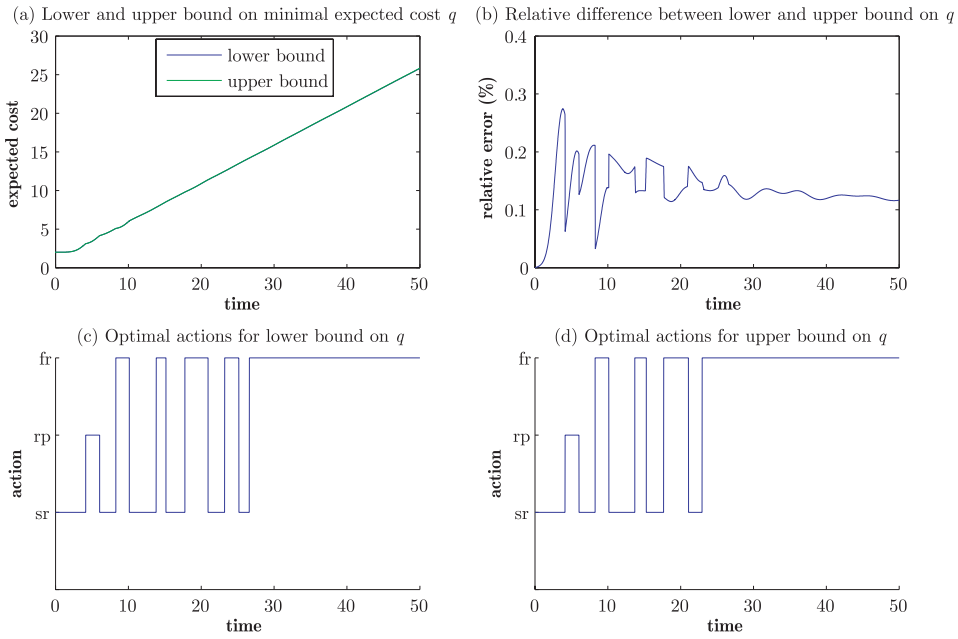

Figure 6.4: (a) Fixed points $\underline{q}$ and $\overline{q}$ of the operators L_h and U_h, respectively. (b) Relative difference $100\big(\overline{q}(t_n) - \underline{q}(t_n)\big)/\overline{q}(t_n)$. (c)-(d) Optimal maintenance actions associated with $\underline{q}$ and $\overline{q}$, respectively.

We are now interested in $q(t)$, the minimal expected maintenance costs $t \in [0, T]$ time units before the end of the horizon, just before a maintenance action has to be selected. Algorithm 6.14 in Appendix 6.B has been utilized to compute $\underline{q}$ and $\overline{q}$ with discretization step $h = 0.01$. It appears from plots (a) and (b) in Figure 6.4 that these bounds on q are approximately equal. That is, in this example the discretization of the dynamic programming equation yields accurate results. Plot (a) also shows that the minimal expected cost $q(t)$ is almost linear in t. Plots (c) and (d) in Figure 6.4 show for both bounds the optimal maintenance action just before the intervention level is exceeded t time units before the end of the horizon. It can be seen that at the end of the horizon 'spot repair' is optimal, whereas 'replacement' is most beneficial on the long term. It is noteworthy that 'repainting' is cost optimal when $t \in [5, 7]$ and that the optimal action changes a number of times from 'spot repair' to 'replacement' for $t \in [10, 25]$. Here the

end-of-horizon effect plays a role: at the end of the horizon the cheapest action will be chosen because it will be the last one. The optimal maintenance actions associated with the bounds $\underline{q}$ and $\overline{q}$ are almost the same, again showing that the discretization yields a good approximation. When discounting is involved the approximations are also good. However, the optimal policy, in both the finite and infinite horizon model, is to perform 'spot repair' when the intervention level is exceeded.

Instead of the uniform distribution other distributions may also be chosen for the reduction due to imperfect maintenance. One of the referees has suggested to relate the probability distributions of the reduction in the affected area just after imperfect maintenance to the physics of corrosion. For instance, the physical simulation model presented in Chapter 2 could be extended to estimate the effect of maintenance on the amount of corrosion. On the other hand, in practice some data on the reduction due to maintenance may be available (as in Meier-Hirmer *et al.*, 2005). In both cases the parameters of some specific distribution can be estimated from the sampled data by using the method of moments. Let us consider two classes of distributions for the random reduction R: (i) a power transformation of a uniform random variable (PTU) on $[0, b]$, defined by bU^a, $a, b > 0$, with U standard uniform, (ii) a shifted Beta random variable on $[a, b]$, $a, b > 0$ fixed, with parameters $\alpha, \beta > 0$. Both random variables are bounded and cover a wide range of density functions. The two parameters of the PTU and shifted Beta distribution can be estimated by means of a two-moment fit (see Tijms, 2003). That is, the sample mean m and sample variance s^2 are equated to the expectation and variance and next the system of two equations in two unknowns is solved. Instead of the variance the coefficient of variation (CV) of R, defined by

$$CV(R) = \frac{(\mathrm{Var}(R))^{1/2}}{\mathbb{E}(R)}$$

can also be used. For the above two distributions the CV can take any positive value.

action a	R_a	PTU		Beta on $[5, 10]$	
spot repair	$\mathbb{E}(R_a)$	7.5	7.5	7.5	7.5
	$CV(R_a)$	$\sqrt{3}/9$	0.4	$\sqrt{3}/9$	0.25
				Beta on $[10, 15]$	
repainting	$\mathbb{E}(R_a)$	12.5	12.5	12.5	12.5
	$CV(R_a)$	$\sqrt{3}/15$	0.2	$\sqrt{3}/15$	0.15

Table 6.1: Parameters settings imperfect reduction.

Table 6.1 shows the settings of the mean and coefficient of variation used in the experiments. Note that the PTU and Beta distribution following from the settings in the

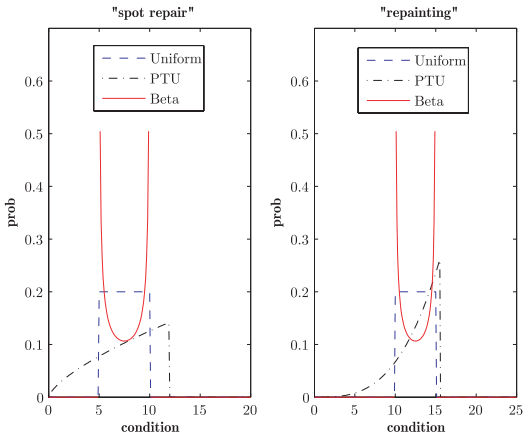

Figure 6.5: Uniform, PTU and shifted Beta density functions.

third and fifth column have the same mean and variance as the uniform random variables in our first example. The settings in the fourth and sixth column yield distributions with higher variability. The density functions in Figure 6.5 correspond to the uniformly distributed reductions used in our first example and the PTU and shifted Beta distributed reductions in the fourth and sixth column of Table 6.1. Whereas the mass of the uniform density is evenly spread, the PTU density is skewed to the right. Most of the mass of the shifted Beta distribution is in its tails.

Let us summarize the main results of the four experiments. First of all, if the reduction has a PTU distribution, then the computing time needed to evaluate the CDF F_a, $a \in \{a_s, a_r\}$, in 5000 points takes 0.5 seconds (approx) vs 2.5 seconds (exact). If the reduction is shifted beta distributed the absolute gain in computing time is much higher: 5 seconds (approx) vs 140 seconds (exact). Regarding the settings in the third and fifth column of Table 6.1 the outcomes are similar to that of the first experiment, because the reductions have approximately the same distribution function. The settings in the fourth and sixth column (high CV) yield very different distributions and hence the optimal policies are different. Figure 6.6 and 6.7 show the optimal expected (non-discounted) cost q, the discretization error and the optimal actions following from the upper and lower bound on q for the Beta distributed reductions. In the PTU case the action 'spot repair' is only optimal at the end of the horizon (i.e. t small), whereas in the Beta case it is also optimal for larger values of t. This is because in the PTU case the reduction can also take small values and hence it does not increase the time until next maintenance as much as in the Beta case.

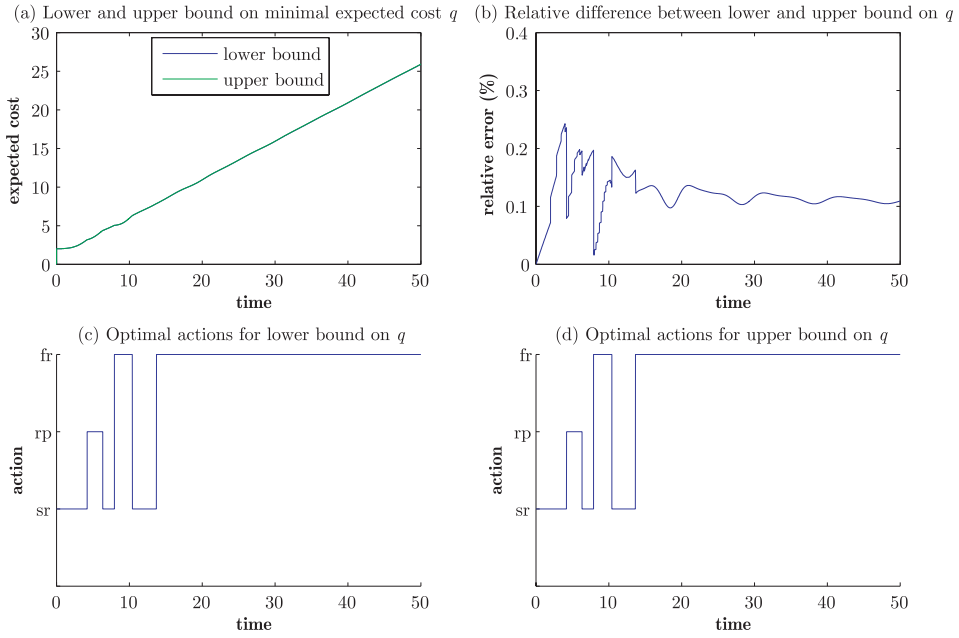

Figure 6.6: Case PTU: (a) Fixed points $\underline{q}$ and $\overline{q}$ of the operators L_h and U_h, respectively. (b) Relative difference $100\big(\overline{q}(t_n) - \underline{q}(t_n)\big)/\overline{q}(t_n)$. (c)-(d) Optimal maintenance actions associated with $\underline{q}$ and $\overline{q}$, respectively.

6.6 Conclusions

The life-cycle management of steel structures involves decisions regarding the timing and the type of maintenance of protective coatings. We have presented a model for optimal maintenance of such coatings on steel structures. The deterioration of coatings is represented by the size of the area affected by corrosion and this is modelled by a non-stationary gamma process. Imperfect maintenance actions such as spot repair and re-painting reduce this size by a random amount, whereas replacement reduces the size to zero. After maintenance the size again follows a non-stationary gamma process with possibly different parameters. It is assumed that maintenance is done as soon as the gamma process exceeds a fixed threshold and consequently the time between two maintenance actions is the first time a gamma process exceeds some nonnegative random threshold. The problem is to find the sequence of maintenance actions that minimizes the expected (discounted) costs over a finite horizon. The continuous-time problem is formulated as a renewal-type optimality equation and it is solved by discretizing time. It is shown that the discretization yields an accurate approximation of the original problem. The outcomes of numerical experiments suggest that different maintenance actions can be optimal over the

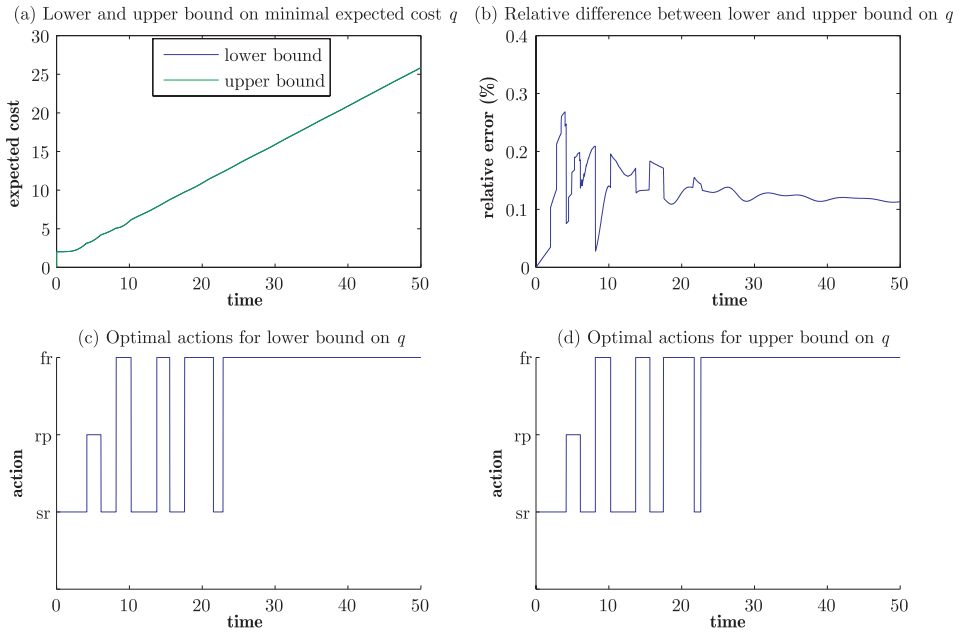

Figure 6.7: Case Beta: (a) Fixed points $\underline{q}$ and $\overline{q}$ of the operators L_h and U_h, respectively. (b) Relative difference $100\big(\overline{q}(t_n) - \underline{q}(t_n)\big)/\overline{q}(t_n)$. (c)-(d) Optimal maintenance actions associated with $\underline{q}$ and $\overline{q}$, respectively.

decision horizon. That is, at the beginning of the horizon the cost and the time until next maintenance (lifetime extension due to the reduction in deterioration) associated with a maintenance action are weighed against another, taking into account that there will be done more maintenance actions in the future. At the end of the horizon the cheapest action is always chosen.

6.A On the overshoot of a non-stationary gamma process

In this appendix we give an upper bound on the expected overshoot $\mathbb{E}(W_{v,\lambda}(r))$ of a non-stationary gamma process for any $r > 0$. By relation (6.4) we know that

$$W_{v,\lambda}(r) \stackrel{d}{=} \lambda^{-1} W(\lambda r)$$

with $W(\lambda r)$ the overshoot at level λr of a standard gamma process and so we only need to determine an upper bound on the expected overshoot of a standard gamma process.

Introduce for any $h > 0$ the hitting time

$$T_h(r) := \min\{nh > 0 : X(nh) > r\} \tag{6.54}$$

with X a standard gamma process. By the definition of the hitting time $T(r)$ of a standard gamma process it is obvious that

$$T(r) \leq T_h(r) \tag{6.55}$$

almost surely. Also, since the standard gamma process has independent and identically distributed increments the random variables $S_n := X(nh)$ form a partial sum process and for this process we introduce the renewal process $N_h := \{N_h(t) : t \geq 0\}$ given by

$$N_h(t) = \sup\{n \in \mathbb{N} \cup \{0\} : S_n \leq t\}. \tag{6.56}$$

This shows by relations (6.54) and (6.56) that

$$T_h(r) = h(N_h(r) + 1)$$

and since X has increasing sample paths we obtain by relation (6.55) that

$$\begin{aligned} W(r) &= X(T(r)) - r \\ &\leq X\big(h(N_h(r) + 1)\big) - r \\ &= S_{N_h(r)+1} - r \end{aligned}$$

This implies using the well-known relation $\mathbb{E}(S_{N_h(r)+1}) = \mathbb{E}(N_h(r) + 1)\mathbb{E}(X(h))$ that

$$\mathbb{E}(W(r)) \leq \mathbb{E}(N_h(r) + 1)\mathbb{E}(X(h)) - r. \tag{6.57}$$

Since $X(h) \sim$gamma$(h, 1)$ we obtain

$$\mathbb{E}(X(h)) = h \text{ and } \mathbb{E}(X^2(h)) = h + h^2.$$

Hence by Lorden's inequality for the renewal function $t \mapsto \mathbb{E}(N_h(t))$ (for an elementary proof of this inequality see Frenk *et al.*, 1997) we obtain

$$\mathbb{E}(N_h(r) + 1) \leq \frac{r}{h} + \frac{h + h^2}{h^2} = \frac{r + 1}{h} + 1.$$

This implies by relation (6.57) that

$$\mathbb{E}(W(r)) \leq h + 1$$

for every $r > 0$. Since this inequality holds for any $h > 0$ we finally obtain that $\mathbb{E}(W(r)) \leq 1$ and so $\mathbb{E}(W_{v,\lambda}(r)) \leq \lambda^{-1}$. This also implies that $\mathbb{E}(W_{v,\lambda}(R)) \leq \lambda^{-1}$ for any random variable R independent of the gamma process. We may assume that in practice the expected reduction $\mathbb{E}(R_a) \gg \lambda^{-1}$ for any action $a \in A$ and so this justifies that we do not consider the overshoot in our model.

6.B Dynamic programming algorithm

The algorithm below simultaneously computes lower and upper bounds for the minimal expected maintenance costs t_n time units before the end of the horizon, just before a maintenance action has to be selected, for $n = 0, 1, \ldots, N$.

Algorithm 6.14
Input: integer N, real T, real array $c(a)$, $\forall a \in A$

INITIALIZATION:
$h = T/N$
compute $p_k^{(a)} := G_a((k+1)h) - G_a(kh)$, $k = 0, 1, \ldots, N-1$, $\forall a \in A$
v(0)=w(0)=0
optactionv(k) = optactionw(k) = 0, $k = 1, 2, \ldots, N$

MAIN:
For $n = 1$ *to* N *do*
  sumv(a) = sumw(a) = 0, $\forall a \in A$
  For $k = 1$ *to* $n-1$ *do*
    For $a \in A$ *do*
      $sumv(a) = sumv(a) + v(n-k)p_{k-1}^{(a)}$
      $sumw(a) = sumw(a) + w(n-k)p_k^{(a)}$
    Next a
  Next k
  $v(n) = \min_{a \in A} \{c(a) + sumv(a)\}$.
  $optactionv(n) = \arg\min_{a \in A} \{c(a) + sumv(a)\}$.
  $w(n) = \min_{a \in A} \left\{(c(a) + sumw(a))/(1 - p_0^{(a)})\right\}$.
  $optactionw(n) = \arg\min_{a \in A} \left\{(c(a) + sumw(a))/(1 - p_0^{(a)})\right\}$.
Next n

OUTPUT:
The arrays $v(n)$, $n = 0, 1, \ldots, N$, and $w(n)$, $n = 0, 1, \ldots, N$, consist of the lower and upper bounds on the minimal expected maintenance costs, t_n time units before the end of the horizon and just before a maintenance action has to be selected. The elements of the arrays optactionv(n), $n = 1, 2, \ldots, N$, and optactionw(n), $n = 1, 2, \ldots, N$, are the maintenance actions to be taken t_n time units before the end of the horizon, just before the intervention level is exceeded.

Chapter 7

A review of multi-component maintenance models*

Abstract

In this chapter we give an overview of the literature on multi-component maintenance optimization. We focus on work appearing since the 1991 survey by Cho and Parlar. Our classification scheme is primarily based on the dependence between components (stochastic, structural or economic). Next, we also classify the papers on the basis of the planning aspect (short-term vs long-term), the grouping of maintenance activities (either grouping preventive or corrective maintenance, or opportunistic grouping) and the optimization approach used (heuristic, policy classes or exact algorithms). Finally, we pay attention to the applications of the models.

7.1 Introduction

Over the last few decades the maintenance of systems has become more and more complex. One reason for this is that systems consist of many components which depend on each other. On the one hand, interactions between components complicate the modelling and optimization of maintenance. On the other hand, interactions also offer the opportunity to group maintenance which may save costs. It follows that planning maintenance actions is a big challenge and it is not surprising that many scholars have studied maintenance optimization problems for multi-component systems. In some articles new solution methods for existing problems are proposed, in other articles new maintenance policies for multi-component systems are studied. Moreover, the number of papers with practical applications of optimal maintenance of multi-component systems is still growing.

*This chapter is based on Nicolai and Dekker (2007b)

Cho and Parlar (1991) give the following definition of multi-component maintenance models: *"Multi-component maintenance models are concerned with optimal maintenance policies for a system consisting of several units of machines or many pieces of equipment, which may or may not depend on each other (economically / stochastically / structurally)."* So, in these models it is all about making an optimal maintenance planning for systems consisting of components which interact with each other. We will come back later to the concepts of optimality and interaction. For now it is important to remember that the condition of the systems depends on (the state of) the components, which will only function if adequate maintenance actions are performed.

Here we will present an up-to-date review of the literature on multi-component maintenance optimization. Let us start with a brief summary of reviews that have appeared in the past. Cho and Parlar (1991) review articles from 1976 up to 1991. The authors divide the literature into five topical categories: machine-interference / repair models, group / block / cannibalization / opportunistic models, inventory / maintenance models, other maintenance / replacement models and inspection / maintenance models. Dekker *et al.* (1996) deals exclusively with multi-component maintenance models that are based on economic dependence. Emphasis is put on articles that have been published after 1991, but there is an overlap with the review of Cho and Parlar (1991). The classification scheme of Dekker *et al.* (1996) differs from that of Cho and Parlar (1991). Firstly, models are classified based on the planning aspect of the model: stationary (long-term) and dynamic (short-term). Secondly, the stationary-grouping models are divided in the categories grouping corrective maintenance, grouping preventive maintenance (PM) and opportunistic grouping maintenance. Here, opportunistic grouping is grouping both preventive and corrective maintenance. The dynamic models are divided into two categories: those with a finite horizon and those with a rolling horizon. Wang (2002) gives an overview of maintenance policies of deteriorating systems. The emphasis is on policies for single component systems. One section is devoted to opportunistic maintenance policies for multi-component systems. The author primarily considers models with economic dependence.

The existing review articles indicate that there are several ways to categorize articles and models. In Section 7.2 we structure the field and present our comprehensive classification scheme. It differs from the schemes used in the review articles discussed earlier. First of all, we distinguish between models with economic, structural and stochastic dependence. Economic dependence implies that grouping maintenance actions either saves costs (economies of scale) or results in higher costs (because of e.g. high downtime costs), as compared to individual maintenance. Stochastic dependence occurs if the condition of components influences the lifetime distribution of other components. Structural dependence applies if components structurally form a part, so that maintenance of a failed

component implies maintenance of working components. In Sections 7.3-7.5, we discuss papers concerning economic, stochastic and structural dependence between components.

In Section 7.6 we classify articles according to the planning aspect of the maintenance model and the method used to optimize the model. Following the review of Dekker *et al.* (1996) we distinguish between models with finite and infinite planning horizons. Models with an infinite planning horizon are called stationary, since they usually provide static rules for maintenance which do not change over the planning horizon. Finite horizon models are called dynamic, since these models can generate dynamic decisions that may change over the planning horizon. In these models short-term information can be taken into account. With respect to the optimization methods, we divide the papers into three categories: exact, heuristic and policy optimization.

Section 7.7 covers trends and open research areas in multi-component maintenance. Conclusions are drawn in Section 7.8.

7.2 Structuring the field

In Section 7.2.1 we give a short review of the terminology used in multi-component maintenance optimization models and explain how we searched the literature. In Section 7.2.2 we present our comprehensive classification scheme.

7.2.1 Search strategy and terminology

Presenting a scientific review on a certain topic implies that one tries to discuss all relevant articles. Finding these articles, however, is very difficult. It depends on the search engines and databases used, electronic availability of articles and the search strategy. We used Google Scholar, Scirus and Scopus as search engines, and ScienceDirect, JStor and MathSciNet as (online) databases. We primarily searched on key words, abstracts and titles, but we also searched within the papers for relevant references. Note that papers published in books or proceedings that are not electronically available, are likely to have not been identified.

Terminology is another important issue, as the use of other terms can hide a very interesting paper. The field has been delineated by maintenance, replacement or inspection on one hand and optimization on the other hand. This combination however, provides almost 5000 hits in Google Scholar.

Next, the term multi-component has been used in junction with related terms as opportunistic maintenance (policies), piggyback(ing), joint replacement, joint overhaul, combining maintenance, grouping maintenance, economies of scale and economic dependence. With respect to the term stochastic dependence, we have also searched for synonyms and related terms such as failure interaction, probabilistic dependence and shock

damage interaction. This yields approximately 500 hits. Relevant articles have been selected from this set by scanning the articles.

The vast literature on maintenance of multi-component systems has been reviewed earlier by others. Therefore, we have also consulted existing reviews and overview articles in this field. Moreover, we have applied a citation search (looking both backwards in time and forwards in time for citations) to all articles found. This citation search is an indirect search method, whereas the above methods are direct methods. The advantage of this method is that one can easily distinguish clusters of related articles.

7.2.2 Classification scheme

First of all, we classify the multi-component maintenance models on the basis of the dependence / interaction between components in the system considered. Thomas (1986) defines three different types of interactions: economic, structural and stochastic dependence.

Simply said, economic dependence implies that the cost of joint maintenance of a group of components does not equal the total cost of individual maintenance of these components. The effect of this dependence comes to the fore in the execution of maintenance activities. On the one hand, the joint execution of maintenance activities can save costs in some cases (e.g. due to economies of scale). On the other hand, grouping maintenance may also lead to higher costs (e.g. due to manpower restrictions) or may not be allowed. For this reason, we will subdivide the models with economic dependence into two categories: positive and negative economic dependence. That is, we refine the definition of economic dependence as compared to the definition used in the review article of Dekker *et al.* (1996). Note that in many systems both positive and negative economic dependence between components are present. We give special attention to the modelling of maintenance optimization of these systems, in particular the k-out-of-n system.

Stochastic dependence occurs if the condition of components influences the lifetime distribution of other components. Synonyms of stochastic dependence are failure interaction or probabilistic dependence. This kind of dependence defines a relationship between components upon failure of a component. For example, it may be the case that the failure of one component induces the failure of other components or causes a shock to other components.

Structural dependence applies if components structurally form a part, so that maintenance of a failed component implies maintenance of working components, or at least dismantling them. So, structural dependence restricts the maintenance manager in his decision on the grouping of maintenance activities.

A second classification of the models is based on the planning aspect: stationary or dynamic. That is, do we make a short-term / operational or a long-term / strategic planning

for the maintenance activities? Is the planning horizon finite or infinite? In stationary models, a long-term stable situation is assumed and mostly these models assume an infinite planning horizon. Models of this kind provide static rules for maintenance which do not change over the planning horizon. They generate for example long-term maintenance frequencies for groups of related activities or control limits for carrying out maintenance depending on the state of components. In dynamic models, short-term information such as a varying deterioration of components or unexpected opportunities can be taken into account. These models generate dynamic decisions that may change over the planning horizon.

The last classification we consider is based on the type of optimization method used. This can be either an exact method, a heuristic or a search within classes of policies. Exact optimization methods are designed to find the real optimal solution of a problem. However, if the computing time of the optimization method increases exponentially with the number of components, then exact methods are only desirable to a certain extent. In that case solving problems with many components is impossible and heuristics should be used. Heuristics are local optimization methods that do not pretend to find the global optimum, but can be applied to find a solution to the problem in reasonable time. The quality of such a solution depends on the problem instance. In some cases it is possible to give an upper bound on the gap between the optimal solution and the solution found by the heuristic.

In many papers a maintenance planning is made by optimizing a certain type of policy. Well known maintenance policies are the age and block replacement policies and their extensions. The advantage of policy optimization over other optimization methods is that it gives more insight into the solution of the problem. Note that policy optimization will not always result in the global optimal solution, since there may be another policy that results in a better solution. In some cases however, it can be proved that applying a certain maintenance policy results in the exact (global) optimal solution.

7.3 Economic dependence

In this section we review articles on multi-component systems with economic dependence. We focus on articles appearing since the review of Dekker *et al.* (1996). In Sections 7.3.1 and 7.3.2 we discuss models with positive and negative dependence, respectively. In Section 7.3.3 we discuss articles on k-out-of-n systems, in which both positive and negative dependence between components are present.

7.3.1 Positive economic dependence

Positive economic dependence implies that costs can be saved when several components are jointly instead of separately maintained. Compared with the review of Dekker *et al.* (1996) we refine the concept of (positive) economic dependence and distinguish the following forms:

- Economies of scale
  - General
  - Single set-up
  - Multiple set-ups
    - ◇ Hierarchy of set-ups
- Downtime opportunity

The term economies of scale is often used to indicate that combining maintenance activities is cheaper than performing maintenance on components separately. The term economies of scale is very general and it seems to be equal to positive economic dependence. Here we shall speak of economies of scale when the maintenance cost per component decreases with the number of maintained components. Economies of scale can result in from preparatory or set-up activities that can be shared when several components are maintained simultaneously. The cost of this set-up work is often called the *set-up cost*. Set-up costs can be saved when maintenance activities on different components are executed simultaneously, since execution of a group of activities requires only one set-up.

In this overview we shall distinguish between single set-ups and multiple set-ups. In the latter case there usually is a hierarchy of set-ups. For instance, consider a system consisting of two components, which both consist of two subcomponents. Maintenance of the subcomponents of the components may require a set-up at system level and component level. Firstly, this means that the set-up cost at component level is paid only once when the maintenance of two subcomponents of a component is combined. Secondly, the set-up cost at system level is paid only once when all subcomponents are maintained at the same time. Set-up costs usually come back in the objective function of the maintenance problem. If economies of scale are not explicitly modelled by including set-up costs in the objective function, then we classify the model in the category 'general'.

Another form of positive dependence is the downtime opportunity. Component failures can often be regarded as opportunities for preventive maintenance of non-failed components. In a series system a component failure results in a non-operating system. In that

case it may be worthwhile to replace other components preventively at the same time. This way the system downtime results in cost savings since more components can be replaced at the same time. Moreover, by grouping corrective and preventive maintenance the downtime can be regulated and in some cases it can even be reduced. Note that if the downtime cost is included in the set-up cost in a certain paper, then we will not classify the paper in the category 'downtime opportunity', but in the category 'set-up cost'. In general however, it is difficult to asses the cost associated with the downtime (see e.g. Smith and Dekker, 1997 who approximate the availability and the cost of downtime for a 1-out-of-n system). Therefore, the downtime cost is usually not included in the set-up cost.

In the paragraphs below we discuss articles dealing with positive economic dependence. Our main focus is on the modelling of this dependence.

Economies of scale

General

In comparison with Dekker *et al.* (1996) the category 'general economies of scale' is new. The papers in this category deal with multi-component systems for which joint maintenance of components is cheaper than individual maintenance of components. This form of economies of scale cannot be modelled by introducing a single set-up cost. The cost associated with the maintenance of components is often concave in the number of components that are maintained simultaneously.

Dekker *et al.* (1998a) evaluate a new maintenance concept for the preservation of highways. In road maintenance cost savings can be realized by maintaining larger sections instead of small patches. The road is divided in sectors of 100 meters length. Set-up costs are present in the form of the direct costs associated with the maintenance of different parts of the road. The set-up cost is a function of the number of these parts in a maintenance group. A heuristic search procedure is proposed to find the optimal maintenance planning.

Hopp and Kuo (1998) consider the maintenance of aircraft engine components. Economies of scale exist (i) for joint replacement because the aircraft must be pulled from service for maintenance and (ii) repair of some components requires removal and disassembly of the engine. Because the joint replacement problem is difficult to solve exactly, the authors introduce heuristics and test their performance. They find that a simple heuristic in which replacement cycles for all components are restricted to be multiples of specified interval is robustly accurate.

Papadakis and Kleindorfer (2005) introduce the concept of network topology dependencies (NTD) for infrastructure networks. In these networks two types of NTD can be distinguished: contiguity and set-up discounts. Both types define positive economic dependence between components. In the former case savings are realized when costs are

paid once when contiguous sections are maintained at the same time. In the latter case savings are realized when costs may be paid once for a neighbourhood of the infrastructure network, independently of how much work is carried out on it. For both types of dependencies a non-linear discount function is defined. The authors consider the problem of maintaining an infrastructure network. It is modelled as an undirected network. Risk measures or failure probabilities for the segments of this network are assumed to be known. A maximum-flow minimum-cut formulation of the problem is developed. This formulation makes it easier to solve the problem exact and efficiently.

Single set-up
Nearly all articles reviewed by Dekker *et al.* (1996) can be classified in this category. The objective function of the maintenance optimization model usually consists of a fixed cost (the set-up cost) and variable costs. In the articles discussed below, this will not be different.

Castanier *et al.* (2005) consider a two-component series system. Economic dependence between the two components is present in the following way. The set-up cost for inspecting or replacing a component is charged only once if the actions on both components are combined. That is, joint maintenance of components saves costs. The condition of the components is modelled by a stochastic process and it is monitored by non-periodic inspections. In the opportunistic maintenance policy several thresholds are defined for doing inspections, corrective and preventive replacements, and opportunistic maintenance. These thresholds are decision variables. Many articles on this type of models have appeared, but most of these articles only consider single component models.

The articles of Scarf and Deara (1998, 2003) consider both economic and stochastic dependence between components in a series system. This combination is scarce in the literature. Positive economic dependence is modelled on the basis that the cost of replacement of one or more components includes a one-off set-up cost whose magnitude does not depend on the number of components replaced. We will discuss these articles in more detail in Section 7.4.

In one of the few case studies found in the literature, Van der Duyn Schouten *et al.* (1998) investigate the problem of replacing light bulbs in traffic control signals. Each installation consists of three compartments for the green, red, and yellow lights. Maintenance of light bulbs means replacement, either correctively or preventively. Firstly, positive economic dependence is present in the form of set-up cost, because each replacement action requires a fixed cost in the form of transportation of manpower and equipment. Secondly, the failure of individual bulbs are opportunities for preventive maintenance on other bulbs. The authors propose two types of maintenance policies. In the first policy, also known as the standard indirect-grouping strategy (introduced in maintenance by Goyal and Kusy, 1985, for a review of this strategy we refer to Dekker *et al.*, 1996),

corrective and preventive replacements are strictly separated. Economies of scale can thus only be achieved by combining preventive replacements of the bulbs. The authors also propose the following opportunistic age-based grouping policy. Upon failure of a light bulb, the failed bulbs and all other bulbs older than a certain age are replaced.

Budai *et al.* (2006) consider a preventive maintenance scheduling problem (PMSP) for a railway system. In this problem (short) routine activities and (long) unique projects for one track have to be scheduled in a certain period. To reduce costs and inconvenience for the travellers and operators, these activities should be scheduled together as much as possible. With respect to the latter, maintenance of different components of one track simultaneously requires only one track possession. Time is discretized and the PMSP is written as a mixed-integer linear programming model. Positive dependence is taken into account by the objective function, which is the sum of the total track possession cost and the maintenance cost over a finite horizon. To reduce possible end-of-horizon effects an end-of-horizon valuation is also incorporated in the objective function. Note that the possession cost can be seen as a downtime cost. The cost is modelled as a fixed / set-up cost. This is the reason that it is classified in this category. Besides this positive dependence there also exists negative dependence between components, since some activities exclude each other. The advantage of a discrete time model is that negative dependence can be incorporated in the model by adding additional restrictions. It appears that the PMSP is a NP hard problem. Heuristics are proposed to find near-optimal solutions in reasonable time.

Multiple set-ups
This is also a new category. The maintenance of different components may require different set-up activities. These set-up activities may be combined when several components are maintained at the same time. We have found one article in this category; it assumes a complex hierarchical set-up structure.

Hierarchical structure of set-ups Van Dijkhuizen (2000) studies the problem of clustering preventive maintenance jobs in a multiple set-up multi-component production system. As far as the authors know, this is the first attempt to model a maintenance problem with a hierarchical (tree-like) set-up structure. Different set-up activities have to be done at different levels in the production system before maintenance can be done. Each component is maintained preventively at an integer multiple of a certain basic interval, which is the same for all components, and corrective maintenance (CM) is carried out in between whenever necessary. So, every component has its own maintenance frequency – the frequencies are based on the optimal maintenance planning for single components. Obviously, set-up activities may be combined when several components are maintained at the same time. The problem is to find the maintenance frequencies that minimize

the average cost per unit of time. This problem is an extension of the standard-indirect grouping problem (for an overview of this problem see Dekker *et al.*, 1996).

Downtime opportunity

As we stated earlier, the downtime of a system is often an opportunity to combine preventive and corrective maintenance. This is specially true for series systems, where a single failure results in a system breakdown. Of course, non-failed components should not be replaced when they are in a good condition, because then useful lifetime can be wasted. The maintenance policies proposed in the articles discussed below use this idea.

Gürler and Kaya (2002) propose a new opportunistic maintenance policy for a series system with identical items. The article is an extension of the work by Van der Duyn Schouten and Vanneste (1993), who also propose an opportunistic policy for such a system. In their model, the lifetime of the components is described by several stages, which are classified as good, doubtful, preventive maintenance due and failed. Gürler and Kaya (2002) classify the stages in the same way, but the stages good and doubtful are subdivided into a number of states. The proposed policy is of the control-limit type. Components which are PM due (failed) are preventively (correctively) replaced immediately. The entire system is replaced when a component is PM due or down and the number of components in doubtful states is at least N. Here, N is a decision variable. It appears that this policy achieves significant savings over a policy where the components are maintained individually without any system replacement.

Popova and Wilson (1999) consider m-failure, T-age and (m,T) failure group policies for a system of identical components operating in parallel. According to these policies the system is replaced at the time of the m-th failure, every T time units, and at the minimum time of these events, respectively. These policies were first introduced by Assaf and Shanthikumar (1987), Okumoto and Elsayed (1983) and Ritchken and Wilson (1990), respectively. Popova and Wilson (1999) assume that downtime costs are incurred when failed components are not repaired or replaced. So, when the system operates there is also negative dependence between the components. After all, when the components are left in a failed condition, with the intention to group corrective maintenance, then downtime costs are incurred. In the maintenance policies a trade-off between the downtime costs and the advantages of grouping (corrective) maintenance is made.

Berk and Moinzadeh (2002) study a similar two-parameter control-limit policy for a system of identical machines where the performance of the machines deteriorates with the operating time since last maintenance. When there are n machines in operation, a machine is taken off line for maintenance if its operating time exceeds a certain threshold T_n. That is, the threshold age depends on the number of machines in operation. Whereas most studies assume the maintenance times are negligible, in this work they are assumed to be independent and identically distributed random variables.

Sheu and Jhang (1996) propose a new two-phase opportunistic maintenance policy for a group of independent identical repairable units. Their model takes into account downtime costs and the maintenance policy includes minimal repair, overhaul, and replacement. In the first interval, $(0, T]$, minor failures are removed by minimal repairs and 'catastrophic' failures by replacements. In the second phase, $(T, T + W]$, minor failures are also removed by minimal repairs, but 'catastrophic' failures are left idle. A group maintenance is conducted at time $T + W$ or upon the kth idle, whichever comes first. The generalized group maintenance policy requires inspection at either the fixed time $T + W$ or the time when exactly k units are left idle, whichever comes first. At an inspection, all idle components are replaced with new ones and all operating components are overhauled so that they become as good as new.

Higgins (1998) studies the problem of scheduling railway track maintenance activities and crews. Positive economic dependence is present in the following way. The occupancy of track segments due to maintenance prevents all train movements on those segments. The costs associated with this can be regarded as downtime costs. The maintenance scheduling problem is modelled as a large scale $0 - 1$ programming problem with many (non-linear) restrictions. The objective is to minimize expected interference delay with the train schedule and prioritized finishing time. The downtime costs are modelled by including downtime probabilities in the objective function. The author proposes tabu search to solve the problem. The neighbourhood, which plays a prominent role in local search techniques, is easily defined by swapping the order of activities or maintenance crews.

The article of Sriskandarajah *et al.* (1998) discusses the maintenance scheduling of rolling stock. Multiple train units have to be overhauled before a certain due date. The aim is to find a suitable common due date for each train so that the due dates of individual units do no deviate too much from the common due date. Maintenance carried out too early or too late is costly since this may cause loss of use of a train. A genetic algorithm is proposed to solve this scheduling problem.

7.3.2 Negative economic dependence

Negative economic dependence between components occurs when maintaining components simultaneously is more expensive than maintaining components individually. There can be several reasons for this.

- Manpower restrictions
- Safety requirements

- Redundancy / production-loss

Firstly, grouping maintenance results in a peak in manpower needs. Manpower restrictions may even be violated and additional labour needs to be hired, which is costly. The problem here is to find the balance between workload fluctuation and grouping maintenance.

Secondly, there are often restrictions on the use of equipment, when executing maintenance activities simultaneously. For instance, use of equipment may hamper use of other equipment and cause unsafe operations. Legal and/or safety requirements often prohibit joint operation.

Thirdly, joint (corrective) maintenance of components in systems in which some kind of redundancy is available may not be beneficial. Although there may exist economies of scale through simultaneous repair of a number of (identical) components, leaving components in a failed condition for some time increases the risk of costly production losses. We will come back to this in Section 7.3.3. Production loss may increase more than linearly with the number of components out of operation. For an example of this type of economic dependence we refer to Stengos and Thomas (1980). The authors give an example of the maintenance of blast furnaces. The disturbance due to maintenance is substantially more, the more furnaces that are out of operation. That is, the cost of overhauling the furnaces increases more than linearly with the number of furnaces out of action.

It appears that maintenance of systems with negative dependence is often modelled in discrete time. The models can be regarded as scheduling problems with many restrictions. These restrictions can easily be incorporated in discrete time models such as (mixed) integer programming models. With respect to these models, there is always the question whether the exact solution can be found efficiently. In other words, the question arises whether the problem is NP-hard. An example of discrete time modelling is given by the article of Grigoriev *et al.* (2006) who study the so-called periodic maintenance problem (PMP). Machines have to be serviced regularly to prevent costly production losses. The failures causing these production losses are not modelled. Time is discretized into unit-length periods. In each period at most one machine can be serviced. Apparently, negative economic dependence in the form of manpower restrictions or safety measures play a role in the maintenance of the machines. The problem is to find a cyclic maintenance schedule of a given length T that minimizes total service and operating costs. The operating costs of a machine increase linearly with the number of periods elapsed since last servicing that machine. PMP appears to be a NP-hard problem and the authors propose a number of solution methods. This leads to the first exact solutions for larger sized problems.

In Stengos and Thomas (1980) time is also discretized but the maintenance problem, scheduling the overhaul of two pieces of equipment, is set up as a Markov decision process. The pieces can be in different states and the probability of failure increases with the time

since the last overhaul. So in comparison with the problem of Grigoriev *et al.* (2006), pieces can fail during operation. Negative economic dependence is modelled as follows. The cost of overhauling the pieces increases more than linearly with the number of pieces out of action. The objective is to minimize the 'loss of production' cost, which is incurred when a piece is overhauled. The optimal policy is found by a relative value successive approximation algorithm.

In Langdon and Treleaven (1997) the problem of scheduling maintenance for electrical power transmission networks is studied. There is negative economic dependence in the network due to redundancy / production-loss. Grouping certain maintenance activities in the network may prevent a cheap electricity generator from running, so requiring a more expensive generator to be run in its place. That is, some parts of the network should not be maintained simultaneously. These exclusions are modelled by adding restrictions to the MIP formulation of the problem. The authors propose several genetic algorithms and other heuristics to solve the problem.

7.3.3 k-out-of-n systems

We shall now discuss the different dependencies in the k-out-of-n system in more detail. This system is a typical example of a system with both positive and negative economic dependence between components. A k-out-of-n system functions if at least k components function. If $k = 1$, then it is a parallel system; if $k = n$, then it is a series system. Let us for the moment distinguish between the cases $k = n$ and $k < n$.

In the series system ($k = n$), there is positive economic dependence due to downtime opportunities. The failure of one component results in an expensive downtime of the system and this time can be used to group preventive and corrective maintenance. Negative economic dependence is not explicitly present in the series system.

If $k < n$, then there is redundancy in the system and it fails less often than its individual components. This way a certain reliability can be guaranteed. Typically, the components of this system are identical which allows for economies of scale in the execution of maintenance activities. It is not only possible to obtain savings by grouping preventive maintenance, but also by grouping corrective maintenance. Note that the latter form of grouping is not advantageous in series systems. In other words, the redundant components introduce additional positive dependence in the system. Whereas positive economic dependence is present upon failure of a component, negative economic dependence plays a role as long as the system operates. A single failure of a component may not always be an opportunity to combine maintenance activities. Firstly, grouping corrective and preventive maintenance upon the failure of the component, increases the probability of system failure and costly production losses. Secondly, leaving components in a failed condition for some time, with the intention to group corrective maintenance at a later

stage, has the same effect. So, there is a trade-off between the potential loss resulting from a system failure and the benefit of joint maintenance.

One problem of optimizing (age-based) maintenance in k-out-of-n systems is the determination of downtime costs, as a failure does not directly result in system failure. Smith and Dekker (1997) derive the uptime, downtime and costs of maintenance in a 1-out-of-n system (with cold standby), but in general it is very difficult to assess the availability and the downtime costs of a k-out-of-n system. Smith and Dekker (1997) optimize the following age-replacement policy. A component is taken out for preventive maintenance and replaced by a stand-by one, if its age has reached a certain value T_{pm}. Moreover, they determine the number of redundant components needed in the system. Chelbi *et al.* (2007) propose a periodic replacement policy to minimize the unavailability of a series system. This policy prescribes joint preventive replacement of all components at a fixed interval; failed components are immediately replaced.

In the maintenance policies considered in the articles below, an attempt is made to balance the negative aspects of downtime costs and the positive aspects of grouping (corrective) maintenance. The opportunistic maintenance policies proposed in these articles are age-based and also contain a threshold for the number of failures (except for the policy introduced by Sheu and Kuo (1994).

In Dekker *et al.* (1998b) the maintenance of light-standards is studied. A light-standard consists of n independent and identical lamps screwed on a lamp assembly. To guarantee a minimum luminance, the lamps are replaced if the number of failed lamps reaches a prespecified number m. In order to replace the lamps the assembly has to be lowered. This set-up activity is an opportunity to combine corrective and preventive maintenance. Several opportunistic age-based variants of the m-failure group replacement policy (in its original form only corrective maintenance is grouped) are considered. Simulation optimization is used to determine the optimal opportunistic age threshold.

Pham and Wang (2000) introduce imperfect preventive maintenance and partial failure in a k-out-of-n system. They propose a two-stage opportunistic maintenance policy for the system. In the first stage failures are removed by minimal repair; in the second stage failed components are jointly replaced with operating components when m components have failed, or the entire system is replaced at time T, whichever occurs first. Positive economic dependence is of an opportunistic nature. Joint maintenance requires less time than individual maintenance.

Sheu and Kuo (1994) introduce a general age replacement policy for a k-out-of-n system. Their model includes minimal repair, planned and unplanned replacements, and general random repair costs. The system is replaced whenever it reaches age T. The long-run expected cost rate is obtained. The aim of the paper is to find the optimal

age replacement time T which minimizes the long-run expected cost per unit time of the policy.

The article of Sheu and Liou (1992) will be discussed in Section 7.4, because they assume stochastic dependence between the components of a k-out-of-n system.

7.4 Stochastic dependence

In the survey of Thomas (1986) multi-component maintenance models with stochastic dependence are considered as a separate class of models. In the more recent review articles this is not the case. In Cho and Parlar (1991) some articles dealing with failure interaction are discussed, but the modelling of failure interaction between components is not. In Wang (2002) nothing is said about systems with failure interaction; articles on this kind of systems only appear in the references. Actually, this is the first publication, since the survey of Thomas (1986), to give a comprehensive review of multi-component maintenance models with stochastic dependence. We do not aim to give solely a list of papers that have appeared. Instead, we want to give insight into the different ways of modelling failure interaction between components and explain the implications of certain approaches and assumptions with respect to practical applicability.

Stochastic dependence, also referred to as failure interaction or probabilistic dependence, implies that the state of components can influence the state of the other components. Here, the state can be given by the age, the failure rate, state of failure or another condition measure. In their seminal work on stochastic dependence, Murthy and Nguyen (1985b) introduce three different types of failure interaction in a *two-component* system.

Type I failure interaction implies that the failure of a component can induce a failure of the other component with a certain probability p (q), and has no effect on the other component with probability $1-p$ ($1-q$). It follows that there are two types of failures: natural and induced. The natural failures are modelled by random variables, the induced failures are characterized by the probabilities p and q. In Murthy and Nguyen (1985a) the authors extend type I failure interaction to systems with multiple components. It is assumed that whenever a component fails it induces a total failure of the system with probability p and has no effect on the other components with probability $(1-p)$. Here we shall consider this to be the definition of type I failure interaction.

Type II failure interaction in a two-component system is defined as follows. The failure of component 2 can induce a failure of component 1 with probability q, whereas every failure of component 1 acts as a shock to component 2, without inducing an instantaneous failure, but affecting its failure rate.

Type III failure interaction implies that the failure of each component affects the failure rate of the other component. That is, every failure of one of the components acts as a shock to the other component.

A potential problem of the failure rate interaction defined by the last two types, is determining the size of the shock. In practice it is very difficult to assess the effect of a failure of one component on the failure rate of another component. Usually there is not much data on the course of the failure rate of a component after the occurrence of a shock. Shocks can also be modelled by adding a (random) amount of damage to the state of another component. Natural failures then occur if the state of a component (measured by the cumulative damage) exceeds a certain level. We bring this modelling of type II and III failure interaction together in one definition. That is, we renew the definition of type II failure interaction for multi-component systems. It reads as follows: *The system consists of several components and the failure of a component affects either the failure rate of or causes a (random) amount of damage to the state of one or more of the remaining components.* It follows that we regard a mixture of induced failures and shock damage as type II failure interaction. Models with type II failure interaction will also be called shock damage models.

In general, the maintenance policies considered in the literature on stochastic dependence, are mainly of an opportunistic nature, since the failure of one component is potential harmful for the other component(s). Modelling failure interaction appears to be quite elaborate. Therefore, most articles only consider two-component systems. Below we review the articles on failure interaction in the following order. Firstly, we will discuss the type I interaction models. For this type of interaction different opportunistic versions of the well-known age and block replacement policies have been proposed. Secondly, the articles on type II interaction will be reviewed. We will see that in most of these articles the occurrence of shocks is modelled as a non-homogeneous Poisson process (NHPP) or that the failure rate of components is adjusted upon failure of other components. Thirdly, we pay attention to articles that consider both types of failure interaction. Finally, we discuss other forms of modelling failure interaction.

Type I

Murthy and Nguyen (1985a) consider two maintenance policies in a multi-component system with type I failure interaction. Under the first policy all failed components are replaced by new ones. When there is no total system failure, then only the single failed component is replaced. Under the second policy all components, also the functioning component(s), are replaced. When there is no total system failure, then the single failed component is subjected to minimal repair and made operational. The failure rate of the failed component after repair is the same as that just before failure. The authors deduce both the expected cost of keeping the system operational for a finite time period as well

as the expected cost per unit time, of keeping the system operational for an infinite time period.

Sheu and Liou (1992) consider an optimal replacement policy for a k-out-of-n system subject to shocks. Shocks arrive according to a NHPP. The system is replaced preventively whenever it reaches age $T > 0$ at a fixed cost c_0. If the m-th shock arrives at age $S_m < T$, it can cause the simultaneous failure of j components at the same time with probability $p_j(S_m)$ for $j = 0, 1, 2, .., n$, where $\sum_{j=0}^{n} p_j(S_m) = 1$. If $j \geq k$, then the k-out-of-n system is replaced by a new one at a cost c_∞ (unplanned failure replacement). So, the downtime is used to replace all components. If $0 \leq j < k$, then the system is minimally repaired with cost $c_j(S_m)$. After a complete replacement (either a planned or a failure replacement), the shock process is set to zero. All failures subject to shocks are assumed to be instantly detected and repaired. The aim of the paper is to find the optimal T which minimizes the long run expected cost per unit time of the maintenance policy.

The articles of Scarf and Deara (1998, 2003) consider failure-based, (opportunistic) age, and (opportunistic) block replacement policies for a labelled two-component *series* system with type I failure interaction. The articles can be seen as an extension of the article of Murthy and Nguyen (1985b) on failure-based replacement for such systems. Note that since we deal with a series system, the failure of either component causes a system downtime. So, if the system is down, this does not necessarily mean that both components have failed. Economic dependence is modelled on the basis that the cost of replacement of one or more components includes a one-off set-up cost whose magnitude does not depend on the number of components replaced.

The maintenance policies considered in Scarf and Deara (1998) are of the age-based replacement type: replace a component on failure or at age T, whichever is sooner. Failure-based maintenance is viewed as the limiting case ($T \to \infty$) of age-based replacement. As there is also economic dependence between components, the authors consider opportunistic age-based replacement policies: replace a component on failure or at age T or at age $T' < T$ if an opportunity exists.

The policies considered in Scarf and Deara (2003) are of the block replacement type and are extended for two-component systems. The independent block replacement policy is a single component policy and it is of the following form: replace all failed components, replace component 1 at times $k\Delta_1, k = 1, 2, ...$ and replace component 2 at times $k\Delta_2, k = 1, 2, ...$ Block replacement can be grouped: replace failed components and replace the system at times $k\Delta, k = 1, 2, ...$ It can also be combined: replace both components (whether failed or not) on failure of the system and replace the system at times $k\Delta, k = 1, 2, ...$ In modified block replacement policies for a two-component system, a component is only replaced at the block replacement times if its age is greater than some critical value. The block replacement times may be independent or grouped, or the components may be combined. Opportunistic modified block replacement policies are of the form: on

failure of component 1, if the age of component 2, τ_2, is greater than b_2', then replace both components; otherwise just replace component 1. On failure of component 2, if the age of component 1, τ_1, is greater than b_1', then replace both components; otherwise just replace component 2. At block replacement times for component 1, $k\Delta_1, k = 1, 2, ...,$ replace component 1 if $\tau_1 > b_1$ and replace component 2 if $\tau_2 > b_2'$; at block replacement times for component 2, $k\Delta_2, k = 1, 2, ...,$ replace component 2 if $\tau_2 > b_2$ and replace component 1 if $\tau_1 > b_1'$ (for suitable chosen thresholds, b_1, b_1', b_2, b_2').

In both articles the maintenance policies are considered in the context of the clutch system used in a bus fleet. This system consists of the clutch assembly (component 2) and the clutch controller (component 1). Actually, the failure of the controller causes a failure of the assembly with probability 1 and the failure of the assembly causes a failure of the controller with probability 0. It is important to mention that the maintenance policies are not only compared on the basis of cost, but also on ease of implementation and system reliability. It is found that an age-based policy is best, but since this implies that components ages have to be monitored, the authors propose to implement a block or modified block policy. Combined modified block replacements seems to be the best alternative for the clutch system under consideration. Combining maintenance of components has the advantage that the system is in general more reliable, although the long run costs per unit time are higher. The economic gains from using a complex policy have to be weighed up against the addition of investment required to implement such policies.

Jhang and Sheu (2000) address the problem of analyzing preventive maintenance policies in a multi-component system with type I failure interaction. The i-th component $1 \leq i \leq N$ has two types of failures. Type 1 failures are minor failures and are rectified through minimal repair. Type 2 failures are catastrophic failures and induce a total failure of the system (i.e. failure of all other components in the system). Type 2 failures are removed by an unplanned / unscheduled replacement of the system. The model takes into account costs for minimal repairs, replacements and preventive maintenance. Generalized age and block replacement policies are proposed. The age replacement policy implies preventive replacement of all components whenever an operating system reaches age T. In the case of a block replacement policy the system is preventively replaced every T years. The expected long-run cost per unit time for each policy is derived and it is discussed how the optimal T can be determined. Various special cases are discussed in detail. Finally, the authors mention the application of their model to the maintenance of mining cables used in hoisting load.

Type II

Satow and Osaki (2003) consider a two-component parallel system. Component 1 is repairable and at failure minimal repair is done. Failures of component 1 occur according to a NHPP. Whenever the component fails it induces a general distributed random amount

of damage to component 2. The damage is additive and component 2 fails whenever the total damage exceeds a certain failure level. A system failure always occurs whenever component 2 fails, because both components fail simultaneously. By assumption component 2 is not repairable. This means that a failed system needs to be replaced by a new one. Since preventive replacement is cheaper than failure replacement, a two-parameter *preventive* replacement policy is analyzed. The policy takes into account both system age and the total damage of component 2. The system is replaced preventively whenever the total damage of component 2 exceeds k or at time T and it is replaced correctively at system failures. An expression for the expected cost per unit time for long run operation is derived and the policy is optimized analytically for two special cases (the one-parameter policies). Numerical examples show that the policy imposing a limit on the total damage (k) of component 2 outperforms the age T policy. It appears that the two-parameter preventive maintenance policy does not necessarily lead to lower expected costs. This is because in this model the state of component 2 is best indicated by the total damage and its age does not provide any additional information.

Zequeira and Bérenguer (2005) study inspection policies for a two-component parallel standby system. The system operates successfully if at least 1 component functions. Failures can be detected only by periodic inspections. The failure times are modelled as independent random variables. Type II failure interaction is modelled as follows. The failure of one component modifies the (conditional) failure probability of the other component with probability p and does not influence the failure time with probability $1 - p$. Within this respect, the model extends the failure rate interaction models proposed by Murthy and Nguyen (1985b). Inspections are either staggered, i.e. the components are inspected one at a time, or non-staggered, i.e. the components are inspected simultaneously at the same time. It is assumed that there are no economies of scale by doing non-staggered inspections. Numerical experiments prove that for the case of constant hazard rates, staggered inspections outperform non-staggered inspections on the expected average cost per unit time criterion. The authors explain this counter-intuitive result as follows. When inspections are staggered, at least one component is in an operating condition more frequently than when inspections are not staggered.

Lai and Chen (2006) consider a two-component system with failure rate interaction. The lifetimes of the components are modelled by random variables with increasing failure rates. Component 1 is repairable and it undergoes minimal repair at failures. That is, component 1 failures occur according to a NHPP. Upon failure of component 1 the failure rate of component 2 is modified (increased). Failures of component 2 induce the failure of component 1 and consequently the failure of the system. The authors propose the following maintenance policy. The system is completely replaced upon failure, or preventively replaced at age T, whichever occurs first. The expected average cost per unit

time is derived and the policy is optimized with respect to parameter T. The optimum turns out to be unique.

Barros *et al.* (2006) introduce imperfect monitoring in a two-component parallel system. It is assumed that the failure of component i is detected with probability $1-p_i$ and is not detected with probability p_i. The components have exponential lifetimes and when a component fails the extra stress is placed on the surviving one for which the failure rate is increased. Moreover, independent shocks occur according to a Poisson process. These shocks correspond to common cause failures and induce a system failure. The following maintenance policy is proposed. Replace the system upon failure (either due to a shock or failure of the components separately), or preventively at time T, whichever occurs first. Assuming that preventive replacement is cheaper, the total expected discounted cost over an unbounded horizon is minimized. Numerical examples show the relevance of taking into account monitoring problems in the maintenance model. The model is applied to a parallel system of electronic components. When one fails, the surviving one is overworked so as keep the delivery rate not affected.

Type I and II failure interaction

Murthy and Nguyen (1985b) derive the expected cost of operating a two-component system with type I or type II failure interaction for both a finite and an infinite time period. They consider a simple, non-opportunistic, maintenance policy. Always replace failed components immediately. This means that the system is only renewed if a natural failure induces a failure of the other component.

Nakagawa and Murthy (1993) elaborate on the ideas of Murthy and Nguyen (1985b). They consider two types of failure interaction between two components. In the first case the failure of component 1 induces a failure of component 2 with a certain probability. In the second case the failure of component 1 causes a random amount of damage to the other component. In the latter case the damage accumulates and the system fails when the total damage exceeds a specified level. Failures of component 1 are modelled as a NHPP with increasing intensity function. The following maintenance policy is examined. The system is replaced at failure of component 2 or at the N-th failure of component 1, whichever occurs first. For both models the optimal number of failures before replacing the system as to minimize the expected cost per unit time over an infinite horizon is derived. The maintenance policy for the shock damage model is extended as follows: the system is also replaced at time T. This results in a two-parameter maintenance policy which is also optimized. The authors give an application of their models to the chemical industry; component 1 is a pneumatic pump and component 2 is a metal container. The failure of the pneumatic pump may either lead to an explosion, causing system failure (model 1), or lead to a reduction in the wall thickness of the container (model 2). The extension

of model 2 captures the introduction of preventive maintenance of the container at time T.

Other types

Özekici (1988) considers a reliability system of n components. The state of the system is given by the random vector $\mathbf{X}_t$ of the ages of the components at time t, that is $\mathbf{X}_t = (X_t^1, ..., X_t^n)$. It is assumed that $X_t^i \geq 0$ for all $t > 0$ and $i = 1, ..., N$, where $X_t^i = \infty$ implies that component i is in a failed state at time t. The stochastic structure of the system is that the stochastic process $X = \{X(t),\ t \geq 0\}$ with state-space $[0, \infty)$ is a positive, increasing, right-continuous, and quasi-left continuous, strong Markov process. Stochastic dependence between the components is modelled by making the age (state) of a component at time t dependent on the age of the system up to time t. The failure interaction considered here differs from type I and II failure interaction defined above. It is worth to mention that this paper is written independently of the work of Murthy and Nguyen (1985b; 1985a).

Maintenance is modelled as follows. There are periodic overhauls at which the state of the system is inspected and a replacement decision is made on the components based on the observation of the system. The cost structure of the maintenance decision is very general and consists of two types: costs which only depend on the number of replaced components and costs which depend on the state of the system at the time of inspection. Economic dependence between components is 'hidden' in the former costs. Replacing a group of components together is cheaper than replacing the components separately or in smaller subgroup. The optimal replacement problem is formulated as a Markov decision process. The author proposes a very general class of replacement policies, for which the decision to replace a component depends on the age of all components. It appears to be possible to characterize the optimal solution to the replacement problem. Unfortunately, it cannot be proved that there exists a single critical age for the system, which describes the optimal replacement problem. The author provides some intuitive results, e.g. it is not optimal to replace new components and if the age of components that have to be replaced is increased, then the optimal policy does not change. He also gives an important counter-intuitive result: it is not true that more components are replaced as the system gets older.

Cui and Li (2006) introduce a cumulative damage shock model for a system of components with a general opportunistic repair policy. Any arriving shock simultaneously inflicts random damages on al the components in the system. A component fails when its cumulative damage exceeds a given threshold. Any such a failure creates a maintenance opportunity and triggers a simultaneous repair on all the components, including the failed ones, such that damages accumulated at various components are reduced to a certain degree (imperfect repair). It appears that this shock model experiences fewer failures than a similar model without opportunistic maintenance. Li and Xu (2004) also address

the question "Under what conditions can a maintenance policy stochastically reduce the number of failures in a multivariate repairable system?". They introduce a coordinated random group replacement policy for such a system and find that it is often better to have simultaneous replacements of components, rather than replacing each component independently of the others.

7.5 Structural dependence

Structural dependence means that some operating components have to be replaced, or at least dismantled, before failed components can be replaced or repaired. In other words, structural dependence between components indicates that they cannot be maintained independently. This is not about failure dependence, but maintenance dependence. Since the failure of a component offers an opportunity to replace other components, opportunistic policies are expected to perform well on systems with structural dependence between components. Obviously, preventive maintenance may also be advantageous, since maintenance of structural dependent components can be grouped.

There may be several reasons for structural dependence. For example a bicycle chain and a cassette form a union which should always be replaced together, rather than individually. Another example is from Dekker *et al.* (1998a), which considers road maintenance. Several deterioration mechanisms affect roads, e.g. longitudinal and transversal unevenness, cracking and ravelling. For each mechanism one may define a virtual component, but if one applies a maintenance action to such a component it also affects the state with respect to the other failure mechanisms. In Dekker *et al.* (1998a) the effect of a maintenance action on each of the mechanisms is assumed to be known. On the other hand, one may also define a performance indicator to combine the individual deterioration mechanisms to a single summary scale (see e.g. Aven, 1996).

The seminal work in this category is from Sasieni (1956). He considers the production of rubber tyres. The machine which produces the tyres consists of two 'bladders'; one tyre is produced on each bladder simultaneously. Upon failure of a bladder, the machine must be stripped down before replacement can be done. This means that the other bladder can be replaced at the same time. Note that immediate replacement is not mandatory, but a failed bladder will produce faulty tyres. Two maintenance policies are analyzed and optimized. The first one is a preventive maintenance policy. Bladders which have made a predetermined number of tyres (m) without failure are replaced. The second one is an opportunistic version of the first policy. When a machine is stripped to replace one bladder, replace the other bladder if it has produced more than $n \leq m$ tyres.

7.6 Planning horizon and optimization methods

In this section we classify articles on the basis of the planning horizon of the maintenance model and the optimization methods used to solve this model. Actually, these two concepts are related. The majority of the articles reviewed here, assumes an infinite horizon. This assumption facilitates the mathematical analysis; it is often possible to derive analytical expressions for optimal control parameters and the corresponding optimal costs. So, in the category infinite horizon (stationary grouping) models policy optimization is the most popular optimization method. For convenience we shall not review the articles in this category.

Finite-horizon models consider the system in this horizon only, and hence assume implicitly that the system is not used afterwards, unless a so-called *residual value* is incorporated to estimate the industrial value of the system at the end of the horizon. In the article of Budai *et al.* (2006) the so-called end-of-horizon effect is eliminated by adding an additional term to the objective function. This term values the last interval.

The optimization methods applied to finite horizon models are either exact methods or heuristics.[1] Exact methods always find the global optimum solution of a problem. If the complexity of a optimization problem is high and the computing time of the exact method increases exponentially with the size of the problem, then heuristics can be used to find a near-optimal solution in reasonable time.

The scheduling problem studied by Grigoriev *et al.* (2006) appears to be NP-hard. Instead of defining heuristics, the authors choose to work on a relatively fast exact method. Column-generation and a branch-and-price technique are utilized to find the exact solution of larger-sized problems. The problem considered by Papadakis and Kleindorfer (2005) is first modelled as a mixed integer linear programming problem, but it appears that it can also be formulated as a max-flow min-cut problem in an undirected network. For this problem efficient algorithms exist and thus, an exact method is applicable.

Langdon and Treleaven (1997), Sriskandarajah *et al.* (1998), Higgins (1998) and Budai *et al.* (2006) propose heuristics to solve complex scheduling problems. The first two articles utilize genetic algorithms. Higgins (1998) applies tabu search and Budai *et al.* (2006) define different heuristics that are based on intuitive arguments. In all four articles the heuristics perform well; a good solution is found within reasonable time.

[1] Actually, if the maintenance policy is relatively easy, it is sometimes possible to determine the expected maintenance costs over a finite period of time. For instance, Murthy and Nguyen (1985b; 1985a) consider failure-based policies in a system with stochastic dependence and derive an expression for the expected cost of operating the system for a finite time.

7.7 Trends and open areas

In this section we comment on the future research of optimal maintenance of multi-component systems. We first analyze the trends in modelling multi-component maintenance and then discuss the future research areas in this field.

7.7.1 Trends

The last few years several articles have appeared on optimal maintenance of systems with stochastic dependence. In particular, the shock-damage models have received much attention. One explanation for this is that type II failure interaction can be modelled in several ways, whereas there is not much room for extensions in the type I failure model. Another reason is that since the field of stochastic dependence is not very broad yet, it is easy to add a new feature such as minimal repair or imperfect monitoring to an existing model. Thirdly, many existing opportunistic maintenance policies for systems with economic dependence, have not yet been applied to systems with (type II) failure interaction.

Another upcoming field in multi-component maintenance modelling is the class of finite horizon maintenance scheduling problems. Finite horizon models can be regarded as dynamic models, because short-term information can be taken into account. Maintenance scheduling problems are often modelled in discrete time as mixed integer linear programming problems. These problems can be NP-hard and in that case heuristics or local search methods have to be developed in order to solve the problems to near-optimality efficiently. In the last decade tabu search, genetic algorithms and problem specific heuristics have already been applied to maintenance scheduling problems (see Langdon and Treleaven, 1997; Sriskandarajah *et al.*, 1998; Higgins, 1998; Budai *et al.*, 2006). However, there is still need for better local search algorithms.

7.7.2 Open areas

There is scope for more work in the following areas:

Finite horizon models On the one hand, the class of infinite horizon models has been studied extensively in literature. Based on renewal-reward theory many maintenance policies for stationary grouping models have been analyzed. On the other hand, the class of finite horizon models, which includes many maintenance scheduling problems, has never had that much attention. However, maintenance of multi-component systems has to be made operational. Therefore, finite horizon and especially rolling horizon models, which also take short-term into account, have to be developed. In order to solve these models heuris-

tics / local search methods should be further developed. Exact algorithms also need more attention. The article of Grigoriev *et al.* (2006) shows that some scheduling problems of reasonable size can be optimized exactly in a reasonable time.

Case-studies This review shows that case-studies are not represented very well in the field. This is surprising, since maintenance is an applied topic. To our opinion many models are just (mathematical) extensions of existing models and most of the times models are not validated empirically. Case-studies can lead to new models, both in the context of cost structures and dependencies between components.

Modelling multiple set-up activities In this article we have subdivided the category 'economic dependence' into a number of subcategories. It appears that examples of modelling maintenance of systems with multiple set-up activities are scarce. Therefore, this seems to be a promising field for further research. After all, in many production systems complex set-up structures exist.

Structural dependence The field of structural dependence is wide open. To our opinion there have only been published few articles on this topic.

Stochastic dependence Two decades ago Murthy and Nguyen published two articles on the maintenance of systems with stochastic dependence. Although this topic has had much attention since then, most articles still deal with two-component systems. So, there is still a lot of work to do on modelling maintenance of systems with failure interaction consisting of more than two components.

Combination of dependencies We have seen only one example of the combination of structural and economic dependence, i.e. Scarf and Deara (1998; 2003). We have also reviewed some papers with both positive and negative economic dependence. Obviously, the combination of different types of interaction results in difficult optimization models. So, this is also an opportunity for researchers to come up with some new models.

Simulation optimization We have already said that much work has been done on maintenance policies for the class of infinite horizon models. Many maintenance policies are not analytically tractable and simulation is needed to analyze these policies. We observe that the optimization of policies via simulation is often done by using algorithms for deterministic optimization problems. Methods such as simulated annealing, stochastic approximation, perturbation analysis or response surface methodology may be more efficient. For instance, Heidergott and Yuan (2006) use the measure-valued differentiation

(MVD) method to derive an efficient gradient estimator for the long-run average cost associated with age replacement policies for multi-component maintenance problems.

7.8 Conclusions

We have reviewed the literature on optimal maintenance of multi-component maintenance. We first classified articles on the type of dependence between components: economic, stochastic and structural dependence. Subsequently, we subdivided these classes into new categories. For example, we have introduced the categories positive and negative economic dependence. We have paid attention to articles with both forms of interaction. Moreover, we have defined several subcategories in the class of models with positive economic dependence. With respect to articles in the class of stochastic dependence, we are the first to review these articles systematically.

Another classification has been made on the basis of the planning horizon models and optimization methods. We have focussed our attention on the use of heuristics and exact methods in finite horizon models. We have concluded that this is a promising open research area, especially because the infinite horizon models have always got more attention.

We have discussed the trends and the open areas of research reported in the literature on multi-component maintenance. We have observed a shift from infinite horizon models to finite horizon models and from economic to stochastic dependence. This immediately defines the open research areas, which also include topics such as case-studies, modelling combinations of dependencies between components and modelling multiple set-up activities.

Chapter 8

Summary & Conclusions

In this dissertation we have developed mathematical models for the maintenance of deteriorating systems. In particular, we have focused on systems for which the deterioration can be expressed by a continuous-valued measurable quantity indicating or relating to failures. The motivation of the six studies included in this thesis has been the Dynaform project, which aimed at the development of a maintenance management tool for engineering systems. In this final chapter we summarize our findings and answer the research questions posed in Chapter 1.

Let us first discuss the question "How can we best model measurable deterioration?". It is necessary to model the uncertainty associated with the evolution of deterioration, because maintenance decisions are often based on estimates of the deterioration and the lifetime of a system. Grey-box models based on stochastic processes such as Brownian motion (BM) and the gamma process (GP) have always had much attention in the literature and have been applied to model deterioration of various systems (Doksum and Hóyland, 1992; Whitmore and Schenkelberg, 1997; Çinlar *et al.*, 1977; Van Noortwijk and Klatter, 1999). However, a solid comparison between these models has never been made. In this thesis we have compared BM and GP models on aspects such as statistical fit, computational effort, and applicability, both in a theoretical as well as in an empirical setting.

In Chapter 2 the empirical comparison has been made by means of an application to the deterioration of coatings on steel structures. The model parameters have been estimated from inspection data as well as from expert opinions. We have found that the deterioration of coatings can be adequately described by a Brownian motion with a time transformation and a non-stationary gamma process. It follows that the distribution of the deterioration increments is not identical over time. The variability of the coating deterioration appears to be quite low and this explains why the two models do not differ much in this application. The theoretical comparison in Chapter 4 confirms that the

two processes have approximately the same probabilistic characteristics (including hitting times) if the variability of the process is low.

Concerning the efficiency of the two models we conclude that estimating the GP model from the expert opinions is more time-consuming than estimating the BM model. On the other hand, fitting the two models to the inspection data requires the same computational effort.

All by all, both grey-box models are applicable in this case-study, but the GP model has a slightly better overall performance. Actually, from the results in Chapter 4 we conclude that GP models are to preferred over BM models. The most important reason is that if the uncertainty associated with the evolution of deterioration is high, then Brownian motion is not a proper deterioration model, because it can take large negative values.

As the above-mentioned case study deals with the deterioration of systems subject to outdoor weathering conditions, we have also asked the question "Would it not be better to model the underlying physical process?". To answer this question, we have also introduced a new type of deterioration model in Chapter 2, viz. the simulation of a physical process, being a white-box model. It is based on the physical aspects of deterioration and consequently it may give more insight than BM and GP models. Yet, the latter two models allow for an easier analysis and are often sufficient for statistical inference on system deterioration and lifetime, and hence for maintenance optimization purposes. To complete the comparison of deterioration models, we have estimated the white-box model from the two coating deterioration data sets. The fit to the expert opinion data is not that good. We also find that the parameter estimation is much more time-consuming than it is for the grey-box models. Still, the white-box model gives more insight into the physical process, being the combination of initiation and propagation processes.

In this particular case-study we conclude that in a decision support system for maintenance optimization of steel structures subject to coating deterioration, it may be better to use one of the grey-box models for fast parameter estimation and statistical inference. On the other hand, the simulation model is a nice tool to visualize deterioration.

The reason why white-box models are not employed in the analysis of deterioration is that conventional approaches to statistical analysis are not feasible due to the mathematical intractability of these models. Still, white-box models are potentially powerful, because they can capture structural insights in the underlying physical process. Hence, there is a need for a likelihood-based framework for the analysis of such models. In Chapter 3 we have developed such a framework, where white-box models are represented by discrete event systems. Tools such as simulation, density estimation and stochastic approximation are employed to circumvent the mathematical intractability of the likelihood.

The proposed framework is quite general. It consists of two building blocks. The first one captures the modelling of a discrete event system. We formulate such a dynamic system in terms of a marked point process, a stochastic process that consists of events of a certain type (the mark) taking place at certain points of time. A general, but simple, scheme for simulating marked point processes is given. Secondly, in the parameter estimation block, density estimation is applied to approximate the likelihood from the simulated output. Stochastic approximation is used to maximize the 'simulated' likelihood. The likelihood-based parameter estimation introduced in the second block is applicable to all simulation models. A likelihood is the joint density of observations under a model and hence we can apply density estimation to approximate the likelihood. The 'curse of dimensionality' is a problem in density estimation. To avert this problem, we suggest to partition the observations in independent components.

Finally we have illustrated the above methodology with an application to the inspection data that were also used in Chapter 2. The white-box model that is estimated from these data is somewhat different than the simulation of a physical process in that the propagation of corrosion is also considered a stochastic process. The fit of the models, however, is comparable.

As a part of the comparison between BM and GP models, we have studied in Chapter 4 the impact of these deterioration models on maintenance decisions. In order to do this, we have optimized a simple age replacement policy for deteriorating systems, where deterioration is described by either of these models. The models have been defined in such a way that the first two moments of the implied lifetime are equal, where the implied lifetime is defined as the first time the deterioration exceeds the (fixed) intervention level.

Let us first mention that it appears that estimating the parameters of a non-stationary deterioration process from lifetime data, entails some risk. If the variability of the lifetime is high and the expected deterioration is convex, then the variability of the estimated model will be even higher. As a consequence, the gamma process parameters will be very large, whereas the Brownian motion parameter values are such that this process takes negative values. If the parameters are estimated from deterioration measurements, then the parameters of the gamma process will not be that large, but Brownian motion will still take negative values.

For each of the deterioration models the replacement age has been computed under different scenarios (low versus high variability). The main outcome of this study is that the BM model produces a higher replacement age, because this model induces a lower probability of failure. This is again due to the negative increments of the BM model. This leads to the conclusion that the gamma process is the safer option when modelling

deterioration.

In Chapter 5 we have presented two approximation formulas for the probability distribution of the first time a gamma process exceeds a threshold, where the threshold can be a random variable. Such a randomized hitting time often represents the lifetime of a structure in reliability or condition-based maintenance models. As the computation of its distribution is time-consuming, most studies that propose this model do not present numerical results. So, there is a need for fast and reliable approximations. Numerical experiments show that both our approximation formulas are quite accurate when the random threshold is not too small and the variability of the (non-stationary) gamma process is not too high. This is in accordance with the theoretical results derived in this chapter. The first approximation, which comes down to a linear interpolation of the exact hitting time distribution at integer points, is the more efficient one. This approximation has been invoked in Chapter 6 to compute the probability distribution of the time between two imperfect maintenance actions. Finally, we have also derived explicit formulas for the probability distribution of a randomized hitting time in some special cases. For instance, we find that the first time a stationary gamma process exceeds an exponential random variable follows a Weibull distribution.

The study in Chapter 6 answers two questions, viz. "How to model the effect of imperfect maintenance" and "When should we do which maintenance action". Both questions were motivated by the maintenance of coating systems protecting steel structures, where multiple (imperfect) maintenance actions are available to prevent the steel base from corrosion. We have first proposed a model that describes deterioration in the presence of imperfect maintenance actions. In this model the total amount of deterioration is reduced by a random amount when maintenance is done. Put differently, the amount of deterioration just after maintenance is a random variable. Between two maintenance actions the deterioration is modelled by a non-stationary gamma process. Maintenance actions can also have an effect on the parameters of the gamma process. For example, imperfect maintenance may reduce the amount of deterioration, but afterwards it may increase faster (on average) than before maintenance was done. The deterioration model serves as an input for the model that helps us answer when to do which maintenance action.

The maintenance optimization model presented in Chapter 6 is new in that the objective is to find the set of (imperfect) maintenance actions that minimizes the expected (discounted) maintenance costs over a finite horizon. Whereas most maintenance models are only concerned with the timing, we aim at selecting the appropriate actions. In our model the maintenance times are given by the time the deterioration exceeds a fixed intervention level. As we have explained earlier the time between two maintenance actions is a randomized hitting time. The continuous-time optimization problem can be formu-

lated as a renewal-type optimality equation. It is solved numerically by discretizing time. We have shown that this discretization yields an accurate approximation of the original problem (even if we approximate the distribution of the randomized hitting time). The outcomes of numerical experiments suggest that different maintenance actions can be optimal over the decision horizon. That is, at the beginning of the horizon the cost and the reduction in deterioration associated with a maintenance action are balanced, taking into account that there will be done more maintenance actions in the future; at the end of the horizon the cheapest action is chosen.

Finally, in Chapter 7 we have given a new survey of multi-component maintenance models. We have focused on work appearing since the 1991 survey by Cho and Parlar. Our classification scheme is primarily based on the dependence between components, viz. stochastic, structural or economic. We have done this because the dependence between components determines which type of maintenance policy is to be chosen. For instance, we distinguish between systems with positive and negative economic dependence. Whereas in the former case it is beneficial to group maintenance activities (due to economies of scale), it is not in the latter case (due to manpower restrictions, safety requirements or redundancy).

As we have observed a shift from economic to stochastic dependence, we have paid much attention to the field of 'stochastic dependence'. In fact we are the first to give a systematic review of maintenance models for systems with this dependence. Several ways of how the condition of components influences the lifetime distribution of other components have been discussed. We have also observed a shift from infinite horizon models to finite horizon models. Especially, the use of heuristics in complex discrete time models is flourishing. Open research areas include topics such as case-studies, modelling combinations of dependencies between components, modelling multiple set-up activities and optimal maintenance of systems with structural dependence.

Nederlandse samenvatting (Summary in Dutch)

Complexe systemen zoals bruggen, wegen, dijken, dammen en hoogspanningsmasten spelen een belangrijke rol in onze samenleving. Bruggen en dijken maken vervoer van mensen en goederen van A naar B mogelijk, dammen en dijken beschermen laaggelegen gebieden tegen overstromingen en tenslotte voorzien hoogspanningsmasten huizen en bedrijven van elektriciteit. Helaas zijn al deze systemen onderhevig aan veroudering en verslechtert hun conditie na enige tijd, met eventueel falen tot gevolg. Het is duidelijk dat de conditie en de prestaties van het systeem gerelateerd zijn. In het algemeen geldt hoe slechter de conditie, des te minder de prestaties van het systeem zijn. Onderhoudsacties zoals inspecties, reparaties en vervanging dienen uitgevoerd te worden om het systeem te behouden of terug te brengen in een acceptabele, werkende toestand. Immers, de economische schade door slecht functionerende systemen kan enorm hoog zijn, zoals twee recente voorbeelden hieronder duidelijk maken.

De Hollandse Brug, onderdeel van de A6 tussen Amsterdam en Lelystad, werd op 27 april 2007 afgesloten voor zwaar vrachtverkeer, omdat de constructie niet aan de veiligheidseisen voldeed. Zware vrachtwagens konden gaten in het wegdek veroorzaken en dit zou kunnen leiden tot onveilige verkeerssituaties. Het probleem was dat de conditie van de brug slechter was dan verwacht, de werkelijke sterkte onduidelijk was en meer veroudering voorzien werd.

Transportbedrijven in de regio werden zwaar getroffen door de beslissing om de brug te sluiten voor vrachtverkeer en moesten noodgedwongen langere routes kiezen. De totale kosten werden geraamd op 160 000 euro per dag (Terbruggen, 2007). Een alternatieve, betaalde, verbinding per veerboot werd op 17 september 2007 gerealiseerd en de verwachting is dat de brug in juli 2008 weer voor alle verkeer wordt opengesteld.

Een probleem van een geheel andere omvang is het instorten van de I-35W brug over de Mississippi bij Minneapolis op 1 augustus 2007. Al in de jaren 90 werd gerapporteerd dat de conditie van deze brug ontoereikend was (O'Connell *et al.*, 1991) en in 2005 bleek dat de brug mogelijk aan vervanging toe was (Browning, 2007). Passend onderhoud is helaas niet uitgevoerd en de brug is uiteindelijk ingestort. Blijkbaar was er onvoldoende

kennis over de effecten van de veroudering op de sterkte van de brug. Naast de grote economische schade, kwamen er 13 mensen om bij deze tragische ramp.

Bovenstaande voorbeelden benadrukken het belang om de conditie van het systeem regelmatig te meten en het faaltijdstip te voorspellen. Ze leiden bovendien tot vragen als

- Hoe kunnen we het best veroudering modelleren en voorspellen?
- Hoe modelleer je het effect van onderhoud op de conditie van het systeem?
- Wanneer moet welke onderhoudsactie uitgevoerd worden?

Deze vragen kwamen aan de orde in het kader van het Dynaform project (Dynaform consortium, 2002), een samenwerkingsverband tussen de Erasmus Universiteit Rotterdam, Rotterdam Painting Consultants en SemLab. Dit project liep van september 2002 tot augustus 2004 en werd financieel ondersteund door Senter, een onderdeel van het Ministerie voor Economische Zaken. Het project richtte zich op het ontwikkelen van beslissingsondersteunende software voor onderhoud van complexe systemen zoals civiele kunstwerken. Het werk in dit project heeft een aantal open onderzoeksvragen blootgelegd, welke later verder zijn onderzocht en in dit proefschrift zijn opgenomen. Hieronder wordt beschreven welke vragen dat zijn en wat de resultaten van de uitgevoerde studies zijn.

Omdat onderhoudsbeslissingen vaak gebaseerd worden op schattingen van de veroudering, dient de onzekerheid die daarmee gepaard gaat, gemodelleerd te worden. In dit proefschrift is de aandacht in het bijzonder gericht op wiskundige modellen voor onderhoud aan systemen waarvoor de veroudering uitgedrukt kan worden als een meetbare, continue grootheid die gerelateerd is aan de levensduur van deze systemen. De eerste onderzoeksvraag is "Hoe kan meetbare, continue veroudering het beste gemodelleerd worden?". Zogenaamde 'grey-box' modellen gebaseerd op stochastische processen zoals Brownian motion (BM) en het gamma proces (GP) hebben altijd veel aandacht gehad in de literatuur en zijn toegepast om de veroudering van verschillende systemen te modelleren (Doksum and Hóyland, 1992; Whitmore and Schenkelberg, 1997; Çinlar *et al.*, 1977; Van Noortwijk and Klatter, 1999). Echter, een solide vergelijking tussen deze modellen is nooit gemaakt. In dit proefschrift zijn BM en GP modellen op aspecten zoals statistische fit, rekentijd en toepasbaarheid vergeleken, zowel op theoretische als empirische gronden.

In Hoofdstuk 2 is de empirische vergelijking gemaakt door de veroudering van coatings op staalstructuren, zoals de schuiven van de Haringvlietdam, te beschrijven met beide modellen. Zowel inspectiegegevens als de meningen van experts zijn gebruikt om de model parameters te schatten. Het is gebleken dat de veroudering van coatings goed kan worden beschreven door zowel een Brownian motion met een tijdstransformatie als een niet-stationair gamma proces. De kansverdeling van de aangroei in veroudering blijft dus niet hetzelfde over de tijd. De variatie van de veroudering is vrij laag en dit verklaart

waarom de twee modellen niet veel van elkaar verschillen in deze praktische toepassing. De theoretische vergelijking in Hoofdstuk 4 bevestigt dat de twee stochastische processen ongeveer dezelfde probabilistische eigenschappen hebben als de variatie van het proces laag is.

Het schatten van het GP model op basis van expert meningen kost meer rekentijd dan het schatten van het BM model. De schatting van de twee modellen uit inspectiegegevens vergt ongeveer dezelfde rekentijd.

Beide grey-box modellen zijn dus toepasbaar op deze case study, maar het GP model presteert iets beter. Gezien de resultaten in Hoofdstuk 4 concluderen we dat GP modellen de voorkeur verdienen boven BM modellen. De voornaamste reden hiervoor is de volgende. Brownian motion is niet een geschikt verouderingsmodel als de onzekerheid die gepaard gaat met de ontwikkeling van veroudering groot is, omdat het model in dat geval grote negatieve waarden kan aannemen. Door deze eigenschap wordt de overschrijdingskans onderschat.

Omdat de hierboven besproken case study de veroudering van aan de open lucht blootgestelde systemen behandelt, rijst de vraag "Zou het niet beter zijn om het onderliggende fysische proces te modelleren?". Om deze vraag te beantwoorden is in Hoofdstuk 2 een nieuw type verouderingsmodel geïntroduceerd, te weten de simulatie van een fysisch proces, een zogenaamd 'white-box' model. Het is gebaseerd op de fysische aspecten van veroudering en daarom zou het meer inzicht kunnen geven dan BM en GP modellen. Aan de andere kant is de analyse van laatstgenoemde modellen gemakkelijker en vaak voldoende voor statistische gevolgtrekkingen voor de veroudering en levensduur van systemen, en dus ook voor onderhoudsoptimalisatie. De vergelijking van verouderingsmodellen is gecompleteerd met het schatten van het simulatiemodel van het fysische proces op basis van eerdergenoemde gegevens. De fit van het model is niet zo goed voor de expert meningen. Bovendien is de benodigde rekentijd om het model te schatten aanmerkelijk groter dan voor de 'grey-box' modellen. Toch geeft het simulatiemodel meer inzicht in het fysische proces, zijnde de combinatie van initiatie (ontstaan van corrosie) en een propagatie (uitbreiding van corrosie). Bovendien kan het model gevisualiseerd worden.

De reden waarom white-box modellen meestal niet gebruikt worden om veroudering te analyseren is dat conventionele statistische methodes niet mogelijk zijn, aangezien deze modellen wiskundig gezien onhandelbaar zijn. Toch hebben white-box modellen veel potentie omdat zij structurele inzichten in het onderliggende fysische proces kunnen vastleggen. Er is dus een methodologie voor de analyse van deze modellen benodigd. In Hoofdstuk 3 is een dergelijke methodologie ontwikkeld, waarbij white-box modellen gerepresenteerd worden door discrete event systemen. Technieken als simulatie en stochastische approximatie zijn aangewend om wiskundige moeilijkheden te omzeilen.

De ontwikkelde methodologie bestaat uit twee bouwstenen. Ten eerste wordt het discrete event systeem gemodelleerd. Dit is geformuleerd als een stochastisch proces dat bestaat uit gebeurtenissen van een bepaald type en welke op bepaalde tijdstippen plaatsvinden. Een algemeen, maar eenvoudig, schema om dit proces te simuleren is gegeven. In de tweede bouwsteen wordt de methode van meeste waarschijnlijkheid toegepast om de parameters te schatten uit gegevens over de veroudering van een systeem.

Tenslotte is de methodologie geïllustreerd aan de hand van dezelfde inspectiegegevens die ook in Hoofdstuk 2 zijn gebruikt. Het white-box model dat hierbij is geschat verschilt licht van het model in Hoofdstuk 2, de propagatie is nu ook gemodelleerd als een stochastisch proces; de statistische fit is echter vergelijkbaar.

Als onderdeel van de vergelijking tussen BM en GP modellen is in Hoofdstuk 4 de invloed van deze verouderingsmodellen op onderhoudsbeslissingen bestudeerd. Hiertoe is een eenvoudig onderhoudsmodel geoptimaliseerd, waarbij het systeem vervangen wordt als het een bepaalde leeftijd heeft bereikt. Veroudering is beschreven met BM en GP modellen. Deze zijn zo gedefinieerd dat de eerste twee momenten van de levensduur van het systeem gelijk zijn. De levensduur is hierbij het eerste tijdstip waarop de veroudering een vast interventieniveau overschrijdt.

Er moet opgemerkt worden dat het schatten van parameters van een niet-stationair verouderingsproces op basis van levensduurgegevens enigszins riskant is. Als de variatie van de levensduur hoog is en de verwachte *verouderingsgroei* almaar toeneemt in de tijd, dan is de variatie van de veroudering nog veel hoger. Dientengevolge nemen de parameters van het gamma proces zeer grote waarden aan, terwijl Brownian motion negatieve waarden aanneemt. Als de parameters op basis van verouderingsgegevens geschat worden, dan zijn de parameterwaarden van het gamma proces niet zo groot, maar Brownian motion neemt nog steeds negatieve waarden aan.

Voor beide verouderingsmodellen is nu de optimale vervangingsleeftijd bepaald onder verschillende scenario's - lage variatie versus hoge variatie. Het gebruik van Brownian motion resulteert in een hogere vervangingsleeftijd omdat onder dit model de kans op falen kleiner is. Dit wordt veroorzaakt doordat de aangroei van dit proces negatief kan zijn. De conclusie is dat het gamma proces het best gebruikt kan worden als verouderingsmodel.

In Hoofdstuk 5 zijn wiskundige formules gepresenteerd voor de kansverdeling van het eerste tijdstip waarop een gamma proces een bepaald niveau overschrijdt; dit niveau kan zelf ook een kansvariabele zijn. Deze zogenaamde 'randomized hitting time' representeert vaak de levensduur van verouderende systemen in modellen voor risicoanalyse en onderhoud van deze systemen. Omdat de berekening van de verdeling in het algemeen tijdrovend is, zijn numerieke experimenten vaak afwezig in studies waarin dit model gebruikt wordt. Snelle en accurate benaderingen zijn daarom nodig. Numerieke experi-

menten tonen aan dat de gepresenteerde formules erg precies zijn. De eerste formule, die gebaseerd is op een lineaire interpolatie van de exacte verdeling op gehele punten, blijkt het meest efficiënt. Deze benadering is later in Hoofdstuk 6 aangeroepen om de kansverdeling van de tijd tussen twee imperfecte onderhoudsacties uit te rekenen. Tenslotte zijn exacte formules afgeleid voor de kansverdeling van de randomized hitting time in enkele bijzondere gevallen. Zo blijkt het eerste tijdstip waarop een stationair gamma proces een exponentieel verdeelde kansvariabele overschrijdt zelf ook exponentieel verdeeld te zijn.

De studie in Hoofdstuk 6 beantwoordt twee vragen, te weten "Hoe modelleer je het effect van imperfect onderhoud" en "Wanneer dient welke onderhoudsactie uitgevoerd worden?". Imperfecte onderhoudsacties zijn acties die de conditie van een systeem gedeeltelijk herstellen. De motivatie van beide vragen ligt in het onderhoud van coatings op staalstructuren. Meerdere (imperfecte) onderhoudsacties zoals lokaal bijwerken, overschilderen en vervangen van de coating kunnen aangewend worden om het staal tegen corrosie te beschermen. Ten eerste is een model opgesteld dat de veroudering beschrijft in de aanwezigheid van imperfecte onderhoudacties, welke gedaan worden als de veroudering een vast interventieniveau overschrijdt. In dit model reduceren onderhoudsacties de totale veroudering met een random hoeveelheid. De hoeveelheid veroudering juist na onderhoud is dus een kansvariabele. In de periode tussen twee onderhoudsacties volgt de veroudering een niet-stationair gamma proces. Onderhoudsacties kunnen ook een effect hebben op de parameters van dit proces. De hoeveelheid veroudering slinkt in principe door onderhoud te doen, maar nadien zou het verouderingsproces sneller kunnen verlopen. Het verouderingsmodel dient als invoer voor het model dat de vraag beantwoordt wanneer welke actie gedaan moet worden.

Het doel van het model voor onderhoudsoptimalisatie in Hoofdstuk 6 is om die onderhoudsacties te vinden waarvoor de verwachte kosten over een eindige horizon minimaal zijn. Waar het in de meeste modellen gaat om het tijdstip waarop onderhoud wordt uitgevoerd, beoogt dit model de juiste onderhoudsactie te selecteren op het moment dat onderhoud nodig is, dat is als de veroudering een vast interventieniveau overschrijdt. Zoals gezegd is de tijd tussen twee onderhoudsacties een randomized hitting time. Het wiskundige model achter dit onderhoudsprobleem kan opgelost worden door de tijd te discretiseren. Dit blijkt een goede benadering van het werkelijke model te geven. De uitkomsten van numerieke experimenten laten zien dat verschillende onderhoudsacties optimaal kunnen zijn over de beslissingshorizon. Aan het begin van de horizon worden de kosten en de reductie in hoeveelheid veroudering die gepaard gaan met een onderhoudsactie tegen elkaar afgewogen, waarbij in acht wordt genomen dat er nog meer onderhoudsacties in de toekomst gedaan zullen worden. Zoals verwacht is het aan het eind van de horizon het beste om de goedkoopste actie te selecteren.

Tenslotte voorziet Hoofdstuk 7 in een nieuw overzicht van onderhoudsmodellen voor multi-component systemen. Het hoofdstuk richt zich op de literatuur verschenen sinds het overzicht van Cho en Parlar in 1991. Deze is in eerste instantie geclassificeerd op basis van de afhankelijkheid tussen componenten, te weten probabilistische, structurele of economische afhankelijkheid. De reden voor deze classificatie is dat deze afhankelijkheid vaak bepaalt welk type onderhoudsstrategie gekozen wordt. Er wordt bijvoorbeeld onderscheid gemaakt tussen positieve en negatieve economische afhankelijkheid. Waar het in het eerste geval voordelig is om onderhoudsacties te groeperen (wegens schaalvoordelen), is dit minder aantrekkelijk in het tweede geval (wegens veiligheidseisen, redundantie of restricties op arbeidskapitaal). In systemen met probabilistische afhankelijkheid tussen componenten beïnvloedt de conditie (faalkans) van een component mogelijk de conditie (faalkans) van één of meer andere componenten. Daarom wordt in zulke systemen vaak preventief en correctief onderhoud gecombineerd; dit wordt ook wel "opportunistisch groeperen van onderhoud" genoemd. Structurele afhankelijkheid is van toepassing als componenten een geheel vormen, zodat onderhoud van een gefaalde component betekent dat nog functionerende componenten ook onderhouden moeten worden.

Omdat er de laatste jaren relatief veel aandacht in de literatuur is geweest voor het modelleren van probabilistische afhankelijkheid tussen componenten, is dit laatste gebied uitgelicht. Sectie 7.4 is het eerste echte overzicht van onderhoudsmodellen voor systemen met deze vorm van afhankelijkheid. Verschillende manieren waarop de conditie van componenten de levensduurverdeling van andere componenten beïnvloedt worden behandeld. Ook is een verschuiving van modellen met oneindige horizon naar modellen met eindige horizon geconstateerd. Met name het gebruik van heuristieken (benaderingsmethoden) om zulke modellen op te lossen blijkt in trek. Open onderzoeksgebieden omvatten onderwerpen als case studies, het modelleren van combinaties van afhankelijkheden tussen componenten en optimaal onderhoud van systemen met structurele afhankelijkheid.

Bibliography

Abate, J. and W. Whitt (1992), Numerical inversion of probability generating functions, *Operations Research Letters*, **12**, 245–251.

Abdel-Hameed, M. (1975), A gamma wear process, *IEEE Transactions on Reliability*, **24(2)**, 152–153.

Abdel-Hameed, M. (1987), Inspection and maintenance policies of devices subject to deterioration, *Advances in Applied Probability*, **19**, 917–931.

Abramowitz, M. and I. A. Stegun (1972), *Handbook of Mathematical Functions (Ninth Printing)*, Dover Publishers, New York.

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993), *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Andradottir, S. (1996), A scaled stochastic approximation algorithm, *Management Science*, **42(4)**, 475–498.

Assaf, D. and J. G. Shanthikumar (1987), Optimal group maintenance policies with continuous and periodic inspections, *Management Science*, **33**, 1440–1452.

Aven, T. (1996), Optimal replacement of monotone repairable systems, in S. Özekici (ed.), *Reliability and Maintenance of Complex Systems. NATO ASI series, vol.* 154, Springer, Berlin, pp. 224–238.

Bae, S. J., W. Kuo, and P. H. Kvam (2007), Degradation models and implied lifetime distributions, *Reliability Engineering & System Safety*, **92(5)**, 601–608.

Bakker, J. D., H. J. van der Graaf, and J. M. van Noortwijk (1999), Model of lifetime-extending maintenance, in M. C. Forde (ed.), *Proceedings of the 8th International Conference on Structural Faults and Repair*, Engineering Technics Press, Edinburgh.

Barlow, E. R. and F. Proschan (1975), *Statistical Theory of Reliability and Life Testing: Probability Models*, Holt, Rineghart and Winston, New York.

Barros, A., C. Bérenguer, and A. Grall (2006), A maintenance policy for two-unit parallel systems based on imperfect monitoring information, *Reliability Engineering and System Safety*, **91**, 131–136.

Berk, E. and K. Moinzadeh (2002), Analysis of maintenance policies for M machines with deteriorating performance, *IIE Transactions*, **32**, 433–444.

Bertoin, J. (1996), *Lévy Processes*, Cambridge University Press, Cambridge.

Billingsley, P. (1979), *Probability and Measure*, Wiley, New York.

Bremaud, P. (1999), *Markov Chains: Gibbs fields, Monte Carlo Simulation, and Queues*, Springer-Verlag, New York.

Browning, D. (2007), Bridge was rated 'structurally deficient' in 2005, *Star Tribune*, August 6, 2007, http://www.startribune.com/10204/story/1338970.html. Retrieved on August 20, 2007.

Budai, G., D. Huisman, and R. Dekker (2006), Scheduling preventive railway maintenance activities, *Journal of the Operational Research Society*, **57**, 1035–1044.

Castanier, B., C. Bérenguer, and A. Grall (2003), A sequential condition-based repair/replacement policy with non-periodic inspections for a system subject to continuous wear, *Applied Stochastic Models in Business and Industry*, **19(4)**, 327–347.

Castanier, B., A. Grall, and C. Bérenguer (2005), A condition-based maintenance policy with non-periodic inspections for a two-unit series system, *Reliability Engineering & System Safety*, **87**, 109–120.

Chan, V. and W. Q. Meeker (2001), Estimation of degradation-based reliability in outdoor environments, Tech. rep., Iowa State University. Dept. of Statistics.

Chelbi, A., D. Aït-Kadi, and H. Aloui (2007), Availbility optimization for multi-component systems subjected to periodic replacement, in T. Aven and J. E. Vinnem (eds.), *Risk, Reliability and Societal Safety: Proceedings ESREL 2007, 25-27 June, Stavanger, Norway*, Taylor & Francis, London, pp. 1109–114.

Cho, D. I. and M. Parlar (1991), A survey of maintenance models for multi-unit systems, *European Journal of Operational Research*, **51**, 1–23.

Çinlar, E., Z. P. Bažant, and E. Osman (1977), Stochastic process for extrapolating concrete creep, *Journal of the Engineering Mechanics Division*, **103(EM6)**, 1069–1088.

Cui, L. and H. Li (2006), Opportunistic maintenance for multi-component shock models, *Mathematical Methods of Operations Research*, **63**, 493–511.

Cypra, B. (1987), An introduction to the Ising model, *The American Mathematical Monthly*, **94**, 937–954.

Daley, D. J. and D. Vere-Jones (2003), *An Introduction to the Theory of Point Processes. Vol. I*, second edn., Springer-Verlag, New York.

Dekker, R., R. P. Plasmeijer, and J. H. Swart (1998a), Evaluation of a new maintenance concept for the preservation of highways, *IMA Journal of Mathematics Applied in Business and Industry*, **9**, 109–156.

Dekker, R., F. A. van der Duyn Schouten, and R. E. Wildeman (1996), A review of multi-component maintenance models with economic dependence, *Mathematical Methods of Operations Research*, **45**, 411–435.

Dekker, R., J. R. van der Meer, R. P. Plasmeijer, and R. E. Wildeman (1998b), Maintenance of light-standards – a case-study, *Journal of the Operational Research Society*, **49**, 132–143.

Devroye, L. (1986), *Non-uniform Random Variate Generation*, Wiley, New York.

Doksum, K. A. and A. Hóyland (1992), Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution, *Technometrics*, **34**, 74–82.

Dufresne, F., H. U. Gerber, and E. S. W. Shiu (1991), Risk theory with the gamma process., *ASTIN Bulletin*, **21(2)**, 177–192.

Duong, T. and M. L. Hazelton (2005), Cross-validation bandwidth matrices for multivariate kernel density estimation, *Scandinavian Journal of Statistics*, **32**, 485–506.

Dynaform consortium (2002), DYNAFORM: a dynamic platform for maintenance optimization.

Feller, W. (1971), *An Introduction to Probablity Theory and its Applications; volume II*, Wiley, New York.

Frangopol, D. M., M.-J. Kallen, and J. M. van Noortwijk (2004), Probabilistic models for life-cycle performance of deteriorating structures: review and future directions, *Progress in Structural Engineering and Materials*, **6(4)**, 197–212.

Frenk, J. B. G., R. Dekker, and M. Kleijn (1997), A unified treatment of single component replacement models, *Mathematical Methods of Operations Research*, **45**, 437–454.

Frenk, J. B. G. and R. P. Nicolai (2007), Approximating the randomized hitting time distribution of a non-stationary gamma process, Tech. Rep. EI 2007-18, Econometric Institute.

Gardner, M. (1970), Mathematical games: the fantastic combinations of John Conway's new solitaire game "life", *Scientific American*, **223**, 120–123.

Goyal, S. K. and M. I. Kusy (1985), Determining economic maintenance frequency for a family of machines, *Journal of the Operational Research Society*, **36**, 1125–1128.

Grigoriev, A., J. van de Klundert, and F. Spieksma (2006), Modeling and solving the periodic maintenance problem, *European Journal of Operational Research*, **172**, 783–797.

Gürler, Ü. and A. Kaya (2002), A maintenance policy for a system with multi-state components: an approximate solution, *Reliability Engineering & System Safety*, **76**, 117–127.

Harris, T. E. (1974), Contact interacton on a lattice, *Annals of Probability*, **2**, 969–988.

Heidergott, B. and T. Yuan (2006), Gradient estimation for a multi-component maintenance system using measure-valued differentiation, submitted.

Heutink, A., A. van Beek, J. M. van Noortwijk, H. E. Klatter, and A. Barendregt (2004), Environment-friendly maintenance of protective paint systems at lowest costs, in *XXVII FATIPEC Congres; 19-21 April 2004, Aix-en-Provence*, Paris: AFTPVA, pp. 351–364.

Higgins, A. (1998), Scheduling of railway track maintenance activities and crews, *Journal of the Operational Research Society*, **49**, 1026–1033.

Hogg, R. V. and A. T. Craig (1978), *Introduction to Mathematical Statistics*, fourth edn., MacMillan, London.

Hopp, W. J. and Y.-L. Kuo (1998), Heuristics for multicomponent joint replacement: Applications to aircraft engine maintenance, *Naval Research Logistics*, **45**, 435–458.

Ising, E. (1925), Beitrag zur Theorie des Ferromagnetismus, *Zeitschrift für Physik*, **31**, 253–258.

ISO (1998), *ISO/EN 12944: Paints and varnishes – Corrosion protection of steel structures by protective paint systems*, International Organization for Standardization (ISO), Geneva.

Jhang, J. P. and S. H. Sheu (2000), Optimal age and block replacement policies for a multi-component system with failure interaction, *International Journal of Systems Science*, **31**, 593–603.

Kallen, M.-J. (2007), *Markov Processes for Maintenance Optimization of Civil Infrastructure in The Netherlands*, Ph.D. thesis, Delft University of Technology.

Karlin, S. and H. M. Taylor (1975), *A First Course in Stochastic Processes*, San Diego: Academic Press.

Kawata, T. (1972), *Fourier Analysis in Probability Theory*, Academic Press, New York.

Kleinman, N., J. C. Spall, and D. Q. Naiman (1999), Simulation-based optimization with stochastic approximation using common random numbers, *Management Science*, **45**, 1570–1578.

Kong, M. B. and K. S. Park (1997), Optimal replacement of an item subject to cumulative damage under periodic inspections, *Microelectronics Reliability*, **37(3)**, 467–472.

Kreyszig, E. (1978), *Introductory Functional Analysis with Applications*, Wiley, New York.

Lai, M. T. and Y. Chen (2006), Optimal periodic replacement policy for a two-unit system with failure rate interaction, *The International Journal of Advanced Manufacturing and Technology*, **29**.

Langdon, W. B. and P. C. Treleaven (1997), Scheduling maintenance of electrical power transmission networks using genetic programming, in K. Warwick, A. O. Ekwue, and R. Aggarwal (eds.), *Artificial Intelligence Techniques in Power Systems*, chap. 10, Institution of Electrical Engineers, Stevenage, UK, pp. 220–237.

Lawless, J. and M. Crowder (2004), Covariates and random effects in a gamma process model with application to degradation failure, *Lifetime Data Analysis*, **19**, 213–227.

Li, H. and S. H. Xu (2004), On the coordinated random group replacement policy in multivariate repairable systems, *Operations Research*, **52**, 464–477.

Liao, H., E. Elsayed, and L. Chan (2006), Maintenance of continuously monitored degrading systems, *European Journal of Operational Research*, **175(2)**, 821–835.

Meeker, W. K. and L. A. Escobar (1998), *Statistical Methods for Reliability Data*, Wiley, New York.

Meier-Hirmer, C., F. Sourget, and M. Roussignol (2005), Optimising the strategy of track maintenance, in K. Kolowrocki (ed.), *Advances in Safety and Reliability: Proceedings ESREL 2005, 27-30 June, Gdansk, Poland*, Taylor & Francis, London, pp. 1385–1391.

Murthy, D. N. P. and D. G. Nguyen (1985a), Study of a multi-component system with failure interaction, *European Journal of Operational Research*, **21**, 330–338.

Murthy, D. N. P. and D. G. Nguyen (1985b), Study of two-component system with failure interaction, *Naval Research Logistics Quarterly*, **32**, 239–247.

Nakagawa, T. and D. N. P. Murthy (1993), Optimal replacement policies for a two-unit system with failure interactions, *RAIRO Recherche Operationelle/Operations Research*, **27**, 427–438.

Newby, M. J. (1998), Analysing crack growth, *Journal of Aerospace Engineering (Proceedings of the Institution of Mechanical Engineers part G)*, **212**, 157–166.

Newby, M. J. and C. T. Barker (2006), A bivariate process model for maintenance and inspection planning, *International Journal of Pressure Vessels and Piping*, **83(4)**, 270–275.

Newby, M. J. and R. Dagg (2004), Optimal inspection and perfect repair, *IMA Journal of Management Mathematics*, **15**, 175–192.

Nicolai, R. P., G. Budai, R. Dekker, and M. Vreijling (2004), Modelling the deterioration of the coating on steel structures: a comparison of methods, in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, Danvers, pp. 4177–4182.

Nicolai, R. P. and R. Dekker (2007a), A comparison of two non-stationary degradation processes, in T. Aven and J. E. Vinnem (eds.), *Risk, Reliability and Societal Safety: Proceedings ESREL 2007, 25-27 June, Stavanger*, Taylor & Francis, London, pp. 1109–114.

Nicolai, R. P. and R. Dekker (2007b), Optimal maintenance of multi-component systems: a review, in K. A. H. Kobbacy and D. N. P. Murthy (eds.), *Complex System Maintenance Handbook: Blending Theory with Practice (forthcoming)*, Springer Verlag, Berlin.

Nicolai, R. P., R. Dekker, and J. M. van Noortwijk (2007a), A comparison of models for measurable deterioration: an application to coatings on steel structures, *Reliability Engineering & System Safety*, **92(12)**, 1635–1650.

Nicolai, R. P. and J. B. G. Frenk (2006), Modelling imperfect maintenance of coating systems protecting steel structures, in C. Guedes Soares and E. Zio (eds.), *Safety and Reliability for Managing Risk: Proceedings ESREL 2006, 18-22 September, Estoril, Portugal*, Taylor & Francis, London, UK, pp. 563–570.

Nicolai, R. P., J. B. G. Frenk, and R. Dekker (2007b), Modelling and optimizing imperfect maintenance of steel structures, Tech. Rep. EI 2007-24, Econometric Institute.

Nicolai, R. P. and A. J. Koning (2006), A general framework for statistical inference on discrete event systems, Tech. Rep. EI 2006-45, Econometric Institute.

O'Connell, H. M., R. J. Dexter, and P. M. Bergson (1991), Fatigue evaluation of the deck truss of bridge 9340, Tech. Rep. 10, Minnesota Department of Transportation, St. Paul, Minnesota.

Okumoto, K. and E. A. Elsayed (1983), An optimum goup maintenance policy, *Naval Research Logistics Quarterly*, **30**, 667–674.

Özekici, S. (1988), Optimal periodic replacement of multi-component reliability systems, *Operations Research*, **36**, 542–552.

Pandey, M. D., X.-X. Yuan, and J. M. van Noortwijk (2007), The influence of temporal uncertainty of deterioration in life-cycle management of structures, *Structure and Infrastructure Engineering (forthcoming)*.

Papadakis, I. S. and P. R. Kleindorfer (2005), Optimizing infrastructure network maintenance when benefits are interdependent, *OR Spectrum*, **27**, 63–84.

Park, C. and W. J. Padgett (2005), Accelerated degradation models for failure based on geometric Brownian motion and gamma processes, *Lifetime Data Analysis*, **11**, 511–527.

Park, K. S. (1988a), Optimal continuous-wear limit replacement under periodic inspections, *IEEE Transactions on Reliability*, **37(1)**, 97–102.

Park, K. S. (1988b), Optimal wear-limit replacement under periodic inspections, *IEEE Transactions on Reliability*, **37(3)**, 293–294.

Parzen, E. (1962), On the estimation of a probability density function and the mode, *Annals of Mathematical Statistics*, **33**, 1056–1076.

Pham, H. and H. Wang (2000), Optimal (τ, T) opportunistic maintenance of a k-out-of-n:G system with imperfect PM and partial failure, *Naval Research Logistics*, **47**, 223–239.

Popova, E. and J. G. Wilson (1999), Group replacement policies for parallel systems whose components have phase distributed failure times, *Annals of Operations Research*, **91**, 163–189.

Protter, P. (1992), *Stochastic Integration and Differential Equations (A new approach)*, Springer Verlag, New York.

Quantitative Micro Software (2004), Eviews 5 User Guide, http://www.eviews.com/.

R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

Ritchken, P. and J. G. Wilson (1990), (m, T) group maintenance policies, *Management Science*, **36**, 632–639.

Robert, C. P. and G. Casella (1999), *Monte Carlo Statistical Methods*, Springer, New York.

Ross, S. M. (1990), *Simulation*, Academic Press, San Diego.

Rudin, W. (1976), *Principles of Mathematical Analysis*, third edn., McGraw Hill, New York.

Sadegh, P. (1997), Constrained optimization via stochastic approximation with a simultaneous perturbation gradient approximation, *Automatica*, **33(5)**, 889–892.

Sasieni, M. W. (1956), A Markov chain process in industrial replacement, *Operational Research Quarterly*, **7**, 148–155.

Satow, T. and S. Osaki (2003), Optimal replacement policies for a two-unit system with shock damage interaction, *Computers and Mathematics with Applications*, **46**, 1129–1138.

Scarf, P. A. and M. Deara (1998), On the development and application of maintenance policies for a two-component system with failure dependence, *IMA Journal of Mathematics Applied in Business & Industry*, **9**, 91–107.

Scarf, P. A. and M. Deara (2003), Block replacement policies for a two-component system with failure dependence, *Naval Research Logistics*, **50**, 70–87.

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York.

Sheather, S. J. (2004), Density estimation, *Statistical Science*, **19**, 588–597.

Sheather, S. J. and M. C. Jones (1991), A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society, Series B*, **53**, 683–690.

Sheu, S. H. and J. P. Jhang (1996), A generalized group maintenance policy, *European Journal of Operational Research*, **96**, 232–247.

Sheu, S. H. and C. M. Kuo (1994), Optimal age replacement policy with minimal repair and general random repair costs for a multi-unit system, *RAIRO Recherche Operationelle/Operations Research*, **28**, 85–95.

Sheu, S. H. and C. Liou (1992), Optimal replacement of a k-out-of-n system subject to shocks, *Microelectronics Reliability*, **32**, 649–655.

Singpurwalla, N. D. (1995), Survival in dynamic environments, *Statistical Science*, **10**, 86–103.

Smith, M. A. J. and R. Dekker (1997), Preventive maintenance in a 1 out of n system: the uptime, downtime and costs, *European Journal of Operational Research*, **99**, 565–583.

Sobczyk, K. (1987), Stochastic models for fatigue damage of materials, *Advances in Applied Probability*, **30(4)**, 652–673.

Sobczyk, K. and B. F. Spencer Jr (1992), *Random Fatigue: from Data to Theory*, Academic press, San Diego.

Spall, J. C. (1987), A stochastic approximation technique for generating maximum likelihood parameter estimates, in *Proceedings of the 1987 American Control Conference*, Minneapolis, MN, pp. 1161–1167.

Spall, J. C. (2003), *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Wiley, Hoboken, NJ.

Sriskandarajah, C., A. K. S. Jardine, and C. K. Chan (1998), Maintenance scheduling of rolling stock using a genetic algorithm, *Journal of the Operational Research Society*, **49**, 1130–1145.

Stengos, D. and L. C. Thomas (1980), The blast furnaces problem, *European Journal of Operational Research*, **4**, 330–336.

Steutel, F. W. (1971), *Preservation of Infinite Divisibility under Mixing and Related Topics*, Ph.D. thesis, Technological University Twente.

Steutel, F. W. and J. G. F. Thiemann (1990), The gamma process and the Poisson distribution, in *Limit Theorems in Probability Theory and Statistics (Pecs 1989) Colloq. Math. Soc. Janos Bolyai 57*, Elsevier, Amsterdam, pp. 477–490.

Steutel, F. W. and K. van Harn (2004), *Infinite Divisibility and Probability Distributions on the Real Line*, Marcel Dekker, New York.

Stummel, F. and K. Hainer (1980), *Introduction to Numerical Analysis*, Scottisch Academic Press, Edinburgh.

Terbruggen, S. (2007), Verbod voor vrachtverkeer op Hollandse brug voorbarig: Paniek-sluiting (in Dutch), August 10, 2007, *De Ingenieur*, **12**, 26–31.

Thomas, L. C. (1986), A survey of maintenance and replacement models for maintainability and reliability of multi-item systems, *Reliability Engineering*, **16**, 297–309.

Tijms, H. C. (2003), *A First Course in Stochastic Models*, Wiley, Chichester.

Van der Duyn Schouten, F. A., B. van Vlijmen, and S. Vos de Wael (1998), Replacement policies for traffic control signals, *IMA Journal of Mathematics Applied in Business & Industry*, **9**, 325–346.

Van der Duyn Schouten, F. A. and S. G. Vanneste (1993), Two simple control policies for a multicomponent maintenance system, *Operations Research*, **41**, 1125–1136.

Van Dijkhuizen, G. (2000), Maintenance grouping in multi-setup multi-component production systems, in M. Ben-Daya, S. O. Duffuaa, and A. Raouf (eds.), *Maintenance, Modeling and Optimization*, chap. 12, Kluwer Academic Publishers, pp. 283–306.

Van Noortwijk, J. M. (1996), *Optimal Maintenance Decisions for Hydraulic Structures under Isotropic Deterioration*, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Van Noortwijk, J. M. (2007), A survey of the application of gamma processes in maintenance, *Reliability Engineering & System Safety (forthcoming)*, http://dx.doi.org/10.1016/j.ress.2007.03.019.

Van Noortwijk, J. M. and D. M. Frangopol (2004), Two probabilistic life-cycle maintenance models for deteriorating civil infrastructures, *Probabilistic Engineering Mechanics*, **19**, 345–359.

Van Noortwijk, J. M. and H. E. Klatter (1999), Optimal inspection decisions for the block mats of the Eastern-Scheldt barrier, *Reliability Engineering & System Safety*, **65(3)**, 203–211.

Wand, M. P. and M. C. Jones (1995), *Kernel Smoothing*, Chapman and Hall Ltd., London.

Wang, H. (2002), A survey of maintenance policies of deteriorating systems, *European Journal of Operational Research*, **139**, 469–489.

Wang, W., P. A. Scarf, and M. A. J. Smith (2000), On the application of a model of condition-based maintenance, *Journal of the Operational Research Society*, **51(11)**, 1218–1227.

Wang, X. and V. Nair (2005), A class of degradation model based on nonhomogeneous gaussian process, Tech. rep., University of Michigan.

Whitmore, G. A. and F. Schenkelberg (1997), Modelling accelerated degradation data using Wiener diffusion with a time scale transformation, *Lifetime Data Analysis*, **3**, 27–45.

Whittaker, E. T. and G. N. Watson (1958), *Modern Analysis*, fourth edn., Cambridge University Press, London.

Wilks, S. S. (1962), *Mathematical Statistics*, Wiley, New York.

Zequeira, R. I. and C. Bérenguer (2005), On the inspection policy of a two-component parallel system with failure interaction, *Reliability Engineering and System Safety*, **88**, 99–107.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

370. E. REUBEN, *Fairness in the lab: The effects of norm enforcement in economic decisions.*
371. G.J.M. LINDERS, *Intangible barriers to trade: The impact of institutions, culture, and distance on patterns of trade.*
372. A. HOPFENSITZ, *The role of affect in reciprocity and risk taking: Experimental studies of economic behavior.*
373. R.A. SPARROW, *Health, education and economic crisis: Protecting the poor in Indonesia.*
374. M.J. KOETSE, *Determinants of investment behaviour: Methods and applications of meta-analysis.*
375. G. MÜLLER, *On the role of personality traits and social skills in adult economic attainment.*
376. E.H.B. FEIJEN, *The influence of powerful firms on financial markets.*
377. J.W. GROSSER, *Voting in the laboratory.*
378. M.R.E. BRONS, *Meta-analytical studies in transport economics: Methodology and applications.*
379. L.F. HOOGERHEIDE, *Essays on neural network sampling methods and instrumental variables.*
380. M. DE GRAAF-ZIJL, *Economic and social consequences of temporary employment.*
381. O.A.C. VAN HEMERT, *Dynamic investor decisions.*
382. Z. ŠAŠOVOVÁ, *Liking and disliking: The dynamic effects of social networks during a large-scale information system implementation.*
383. P. RODENBURG, *The construction of instruments for measuring unemployment.*
384. M.J. VAN DER LEIJ, *The economics of networks: Theory and empirics.*
385. R. VAN DER NOLL, *Essays on internet and information economics.*
386. V. PANCHENKO, *Nonparametric methods in economics and finance: dependence, causality and prediction.*
387. C.A.S.P. SÁ, *Higher education choice in The Netherlands: The economics of where to go.*
388. J. DELFGAAUW, *Wonderful and woeful work: Incentives, selection, turnover, and workers' motivation.*
389. G. DEBREZION, *Railway impacts on real estate prices.*

390. A.V. HARDIYANTO, *Time series studies on Indonesian rupiah / USD rate 1995 - 2005.*
391. M.I.S.H. MUNANDAR, *Essays on economic integration.*
392. K.G. BERDEN, *On technology, uncertainty and economic growth.*
393. G. VAN DE KUILEN, *The economic measurement of psychological risk attitudes.*
394. E.A. MOOI, *Inter-organizational cooperation, conflict, and change.*
395. A. LLENA NOZAL, *On the dynamics of health, work and socioeconomic status.*
396. P.D.E. DINDO, *Bounded rationality and heterogeneity in economic dynamic models.*
397. D.F. SCHRAGER, *Essays on asset liability modeling.*
398. R. HUANG, *Three essays on the effects of banking regulations.*
399. C.M. VAN MOURIK, *Globalisation and the role of financial accounting information in Japan.*
400. S.M.S.N. MAXIMIANO, *Essays in organizational economics.*
401. W. JANSSENS, *Social capital and cooperation: An impact evaluation of a women's empowerment programme in rural India.*
402. J. VAN DER SLUIS, *Successful entrepreneurship and human capital.*
403. S. DOMINGUEZ MARTINEZ, *Decision making with asymmetric information.*
404. H. SUNARTO, *Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of people's credit bank.*
405. M.Á. DOS REIS PORTELA, *Four essays on education, growth and labour economics.*
406. S.S. FICCO, *Essays on imperfect information-processing in economics.*
407. P.J.P.M. VERSIJP, *Advances in the use of stochastic dominance in asset pricing.*
408. M.R. WILDENBEEST, *Consumer search and oligopolistic pricing: A theoretical and empirical inquiry.*
409. E. GUSTAFSSON-WRIGHT, *Baring the threads: Social capital, vulnerability and the well-being of children in Guatemala.*
410. S. YERGOU-WORKU, *Marriage markets and fertility in South Africa with comparisons to Britain and Sweden.*
411. J.F. SLIJKERMAN, *Financial stability in the EU.*
412. W.A. VAN DEN BERG, *Private equity acquisitions.*
413. Y. CHENG, *Selected topics on nonparametric conditional quantiles and risk theory.*
414. M.D. DE POOTER, *Modeling and forecasting stock return volatility and the term structure of interest rates.*
415. F. RAVAZZOLO, *Forecasting financial time series using model averaging.*
416. M.J.E. KABKI, *Transnationalism, local development and social security: the functioning of support networks in rural Ghana.*
417. M. POPLAWSKI RIBEIRO, *Fiscal policy under rules and restrictions.*
418. S.W. BISSESSUR, *Earnings, quality and earnings management: the role of accounting accruals.*
419. L. RATNOVSKI, *A Random Walk Down the Lombard Street: Essays on Banking.*