

THESIS FOR THE DEGREE OF DOCTOR OF ENGINEERING

Survival Estimation for Opportunistic Maintenance

JOHAN SVENSSON

CHALMERS | GÖTEBORG UNIVERSITY



Chalmers University of Technology
Department of Mathematical Statistics
and
Göteborg University
Department of Mathematical Statistics
SE-412 96 Göteborg, Sweden
Göteborg, Sweden 2007

Survival Estimation for Opportunistic Maintenance
JOHAN SVENSSON
ISSN 978-91-7291-887-942-6

©JOHAN SVENSSON, 2004

Doktorsavhandlingar vid Chalmers Tekniska Högskola
Ny serie nr 2568 ISSN 0346-718X
Department of Mathematical Sciences, Chalmers University of Technology
Department of Mathematical Sciences, Göteborgs University
SE-412 96 Göteborg, Sweden
Telephone +46 (0)31-772 1000
Göteborg, 2007

Survival Estimation for Opportunistic Maintenance
JOHAN SVENSSON
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University

Abstract

The problem of rational maintenance of aircraft engines is studied with respect to the influence of random events. Aircraft engines can be more economically maintained and resources can be saved if the maintenance process is improved. The starting point is an optimization model suggesting what parts in the engine that should be replaced at each maintenance time. The input data is the age of the details in the engine. Statistical models are developed that estimates the remaining life of the components in the engine. The models work with different kinds of data. The first data set only contains times between repairs and is modeled with a non-stationary renewal process and a non-homogeneous Poisson process. With our data the non-stationary renewal process works better. Different repair stations affect the life of the components, which the non-stationary renewal process manages to model. This model also manages the aging component problem in an effective way. However, in this case no aging is present other than substantial degeneration after the first repair. The second dataset contains crack growth data. The remaining life is modeled with a empirical crack growth model. With data directly indicating the condition of the detail a more precise estimate of the remaining life can be made. In order to get an interface with the optimization model the distributions need to be discrete. Four methods to make discretizations are discussed and adapted to suit the model. The methods are compared and the choice concerning the number of points of support is discussed. Finally the consequence of using a narrow scenario tree is commented upon.

Keywords: non-stationary renewal process; non-homogeneous Poisson process; survival; optimal maintenance; discretization; points of support; crack growth

Acknowledgements

The author wants to thank Dragi Anevski, Jacques de Maré at Chalmers and Thomas Svensson at SP. Furthermore I own a debt of gratitude to Fredrik Plym and Maud Österman at Volvo Aero Corporation. The author also thanks Torgny Almgren at Volvo Aero, Niclas Andréasson and Michael Patriksson at Chalmers and Ann-Brith Strömberg at FCC for nice cooperation. This work was financed by NFFP and VAC.

Contents

List of papers	vi
1 Introduction	1
2 The maintenance process	2
3 Survival analysis and Data at VAC	3
3.1 Survival data	3
3.1.1 Survival estimation with one variable	4
3.1.2 Survival estimation with several variables	5
3.1.3 Components and repair	7
3.2 State data	7
3.2.1 Empirical models	8
3.2.2 Mechanistic models	8
3.3 Illustration of life estimation with state and survival data	8
4 Maintenance optimization	10
4.1 Discretization and scenarios	12
4.2 Multistage problems	12
5 Summary of Paper A	13
6 Summary of Paper B	14
7 Summary of Paper C	14
8 Summary of Paper D	15
9 Improvement of data gathering	16
10 Future work	17

List of papers

This thesis includes the following papers

Paper A Torgny Almgren, Niclas Andréasson, Dragi Anevski, Michael Patriksson, Ann-Brith Strömberg and Johan Svensson *Optimization of opportunistic replacement activities*

Paper B Johan Svensson *Two statistical models used on aircraft engine data modeling times between repairs.*

Paper C Johan Svensson *Assessment of residual life based on retarding crack growth*¹.

Paper D Johan Svensson *Discrete approximations of life distributions in optimal replacement.*

¹First part presented in keynote presentation at *International conference residual fatigue life and life extension of in-service structures JIP 2006.*

1 Introduction

Aircraft engines can be more economically maintained and resources can be saved if the maintenance process is improved. Volvo Aero Corporation (VAC) is a company in Trollhättan that, among other things, maintains aircraft engines. Two connected projects were initiated to improve the maintenance process. One project aimed at optimizing the maintenance, that is create a tool that decide what components in the aircraft engine to replace at a given maintenance occurrence. The other project aimed at estimating the survival of one kind of components in the engine with available data from VAC . This thesis describes the work that has been carried out from the life estimation point of view, which also includes an attempt to incorporate the components with the stochastic lives into the maintenance optimization model.

The criterion for optimality is the expected cost to maintain the engine. Input data for the optimization model contains a description of the age of the deterministic details and the stochastic details. A *deterministic detail* is a component that has a predetermined limited time in service which must not be exceeded. Deterministic details are components that are vital for safety. If a deterministic component fails, there is a risk that the engine will stop functioning. In practice the upper limit on the service time means that most deterministic parts will be replaced long before they are even close to failure. A *stochastic detail* is a component that is not deterministic. Stochastic components are allowed to operate in the aircraft engine until they break. If a stochastic component breaks during a flight the engine will still work but with reduced performance.

In the Volvo military engines RM8 and RM12 there are about 50 components that are not expected to last the pre-specified life of the engines. About 75% of them are deterministic and 25% stochastic. More than one third of all cases when the engines are taken to be repaired are unplanned, triggered by the failure of a stochastic component.

A model that describes the properties of the stochastic components is needed. Such models are developed with the type of data that exists at VAC. There are mainly two kinds of data types available, and they are discussed in a later section. Paper B and C suggest models that can be used to model the life of the stochastic components.

To get the stochastic components to interface with the optimization model, approximations need to be made. The optimization model and the type of approximations carried out are discussed in a later section and in paper D. Paper A relates the results obtained in paper B, C and D to the optimization model. Figure 1 illustrates the relationship between the models.

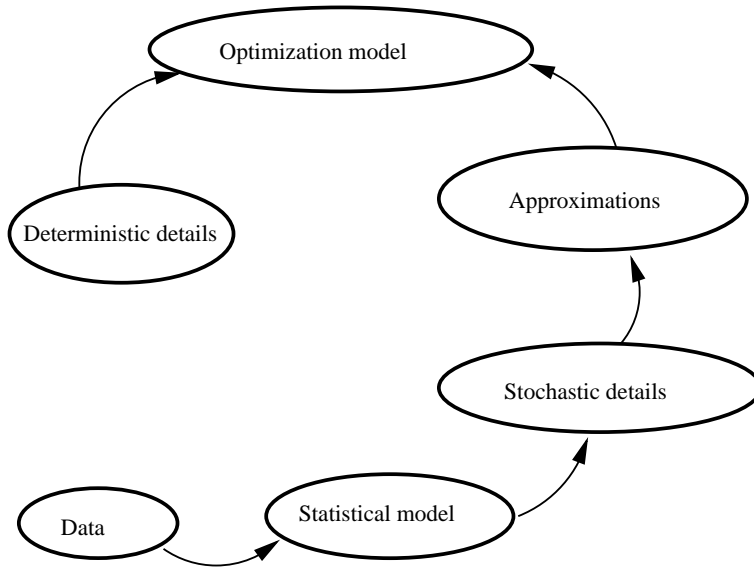


Figure 1: *Illustration of the relationship between the models. The optimization model requires input in the form of the life of the stochastic and deterministic components. The life of the stochastic components needs to be estimated with statistical models that need data to work. Approximations need to be carried out to get the stochastic details to interface with the optimization model.*

2 The maintenance process

There are two principles at VAC which regard to maintenance. The first principle is to maintain the engine at certain fixed predetermined times. The second principle is to maintain the engine only when maintenance is required.

If the engine is maintained according to the first principle the maintenance times are known in advance but not the parts which are to be replaced.

If the maintenance process is ruled by the second principle, the engine is obtained for control only if there are any signs of failure or lacking performance. In this case neither maintenance times nor maintenance needs are known in advance.

The older RM8 military engine was mostly maintained according to the first principle but some components were checked after every flight mission. New construction solutions allow the newer RM12 engine to be examined more carefully on the hangar floor using pipe holes and advanced fiber optics. This opens the possibility for the RM12 engine to be maintained according to the second principle.

When the aircraft lands and the engine is still near the hangar it is examined for

broken components. A component is considered to be broken if it fails to comply with a set of fitness rules when it is observed. If a component is broken the entire engine or possibly a module of the engine is sent to a repair station. Not all components are examined after each flight so in practice this means that the components are checked at certain intervals. These intervals are typically shorter than the predetermined intervals used in the first principle.

If maintenance is carried out according to the second principle there is an opportunity to replace more than the failed components when the engine arrives at the repair station. Such a replacement may decrease the time to next repair and the overall cost of maintaining the engine. An optimization model is developed in order to get suggestions as to what parts in the engine should be replaced.

3 Survival analysis and Data at VAC

The main aim of survival analysis is to determine the life of a population of individuals with the help of data. We will focus on deciding the life for a class of components. At VAC there are mainly two different kinds of data, which we call survival data and state data.

By *survival data* we mean some measure of the life of the component in terms of, e.g., flight hours, cycles or the number of afterburner ignitions. The number of afterburner ignitions may influence the life of the components in the afterburner. A cycle for a civilian aircraft is one takeoff and one landing. An equivalent cycle for a military aircraft engine is a more complex function based on the number of heat and pressure cycles.

By *state data* we mean information related to the current state of the component relevant for the future life. Using the information of the state of the component and a model of the change of the state it is possible to predict when the state of the component is so critical that the component needs to be replaced. Examples of state data are crack length of a component or number of parts burnt away from a component.

With survival data we can answer questions about the population. We can also use the data for estimating the remaining life of a specific component. However, to get a better estimation of the remaining life of a specific component we use state data. The knowledge of the state of a component gives more information about the remaining life than survival data alone, cf. Section 3.3

3.1 Survival data

The life T of the stochastic components at VAC is measured by several variables. Unfortunately all of those variables are highly correlated. This is a consequence of the fact that over longer time periods the engines fly in a similar way (measured in the above variables). This implies that it is not feasible to use regression models, such

as the Cox regression (5) model, to find the effects of covariates in the distribution of component life. In paper B we will therefore limit ourselves to studying only one variable.

3.1.1 Survival estimation with one variable

The methods for estimating survival can be divided into parametric and non-parametric methods. Standard non-parametric estimators are the Nelson-Aalen estimator and the Kaplan-Meier estimator. The Nelson-Aalen estimator estimates the cumulative hazard rate while the Kaplan-Meier estimator estimates the survival function $S(t) = 1 - F(t)$ where $F(t)$ is the cumulative distribution function.

The cumulative hazard rate is defined as

$$H(t) = \int_0^t h(u) du,$$

where

$$h(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log(S(t)) \quad (1)$$

where T is the stochastic variable we wish to describe. The Nelson-Aalen estimator is defined as

$$\hat{H}(t) = \begin{cases} 0 & \text{if } t < t_1 \\ \sum_{i:t_i \leq t} \frac{d_i}{Y(t_i)} & \text{if } t \geq t_1 \end{cases} \quad (2)$$

and the Kaplan-Meier estimator is defined as

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{i:t_i \leq t} (1 - \frac{d_i}{Y(t_i)}) & \text{if } t \geq t_1, \end{cases}$$

where t_i are the observation times, $Y(t_i)$ are the number of individuals at risk just before time t_i , and d_i the number of failures at time t_i .

There exist several parametric distributions that are known to work well for survival data, e.g. the Weibull distribution

$$F(t) = 1 - e^{-(\frac{1}{\theta}t)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0),$$

and the Log-normal distribution

$$F(t) = \Phi\left(\frac{\log t - \mu}{\sigma}\right), \quad t > 0, \quad (\sigma > 0).$$

For parametric models likelihood based estimation approaches are common. The likelihood function for survival data can generally be written as

$$L(\theta) = \prod_{i \in D} f_\theta(t_i) \prod_{j \in R} S_\theta(c_j) \prod_{k \in I} (S_\theta(l_k) - S_\theta(r_k)) \quad (3)$$

where D is the index set for the failure times, R is the index set for the right-censored observations, and I is the set of indices for the interval-censored observations. If individual j is right-censored we know that individual j is alive at time c_j . If individual k is interval-censored we do not know the exact time of the failure but we know that individual k has failed between times l_k and r_k . The parameter vector θ is estimated as

$$\hat{\theta} = \operatorname{argmax}_{\theta} L(\theta). \quad (4)$$

A comparison between the parametric and non-parametric estimates can be made by comparing their hazard rates through hypothesis testing, example one-sample test, cf. Moeschberger and Klein [19].

Inherent in such a comparison is that the parametric distribution is often rejected if the data set is large. The true distribution of the individuals follows no common parametric distribution but has its own distribution. This does not necessarily mean that the parametric distribution is a bad model. This must be taken into account when deciding if the parametric distribution is a good enough model for our application. A visual comparison can be made in several ways. One way is to compare the distribution or density functions between the non-parametric and parametric estimations. An advantage of a visual comparison of the density functions compared to the distribution function is that properties like several humps may be easier to discover which may indicate multiple cause of failures. A non-parametric estimation of the density function can be obtained by performing a kernel smoothing, see further Klein and Moeschberger [19].

3.1.2 Survival estimation with several variables

Assume that we have n observations of the type $(T_j, \delta_j, \mathbf{Z}_j(t))$ where T_j is the time on study for component number j and δ_j is a indicator if the component j has failed ($\delta_j = 1$) or is right censored ($\delta_j = 0$) and $\mathbf{Z}_j(t) = (Z_{j1}, \dots, Z_{jp})$ is a vector of covariates or risk factors. Covariates are variables that are believed to affect the survival of the components. Covariates can be fixed in time, for example an indicator for which country the component is flying in. Different countries affect flying conditions and mission profiles and hence the survival. The covariates can also be changing in time, for example the number of afterburner ignitions. Cox [11] suggests the following semi-parametric proportional hazard model

$$h(t | \mathbf{Z}(t)) = h_0(t)e^{(\beta^T \mathbf{Z}(t))} = h_0(t)e^{(\sum_{k=1}^p \beta_k Z_k(t))}, \quad (5)$$

where $h(t | \mathbf{Z}(t))$ is the hazard rate for a component at risk at time t with covariates $\mathbf{Z}(t)$ and $h_0(t)$ is the baseline hazard. The resulting estimation problem is to estimate the parameter vector β and the baseline hazard $h_0(t)$. Let $R(t)$ be the set of components at risk just before time t . The Cox partial likelihood for the parameter

vector β is then given by

$$L(\beta) = \prod_{i=1}^n \frac{e^{(\sum_{k=1}^p \beta_k Z_{ik}(t_i))}}{\sum_{j \in R(t_i)} e^{(\sum_{k=1}^p \beta_k Z_{jk}(t_i))}}$$

and the estimate is

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} L(\beta).$$

Note that this is not the full likelihood, cf. equation (3). The baseline hazard cannot be estimated directly but the cumulative baseline hazard

$$H_0(t) = \int_0^t h_0(u) du$$

can be estimated with Breslow's estimator

$$\hat{H}_0(t) = \sum_{t_i \leq t} \frac{d_i}{\sum_{j \in R(t_i)} e^{(\hat{\beta}^T \mathbf{Z}_j)}}$$

where t_i are the observation times and d_i the number of failures at time t_i . Note that this estimator reduces to the Nelson-Aalen estimator (2) when there are no covariates present.

An alternative to proportional hazard models is using an additive hazard model. A nonparametric additive hazard model is of the form

$$h(t | \mathbf{Z}(t)) = \beta_0(t) + \sum_{k=1}^p \beta(t)_k Z_k(t),$$

where the parameters $\beta(t) = (\beta_1, \dots, \beta_p)$ are a function of time rather than single parameter values as in the proportional hazard model, cf. Andersen et al. [1]. To be able to compare with the cox regression model the restricted additive model

$$h(t | \mathbf{Z}(t)) = \beta_0(t) + \sum_{k=1}^p \beta_k Z_k(t), \quad (6)$$

can be used. Using a martingale approach (6) can be seen as a regression model, and estimations of the parameters can be obtained by least square estimates, cf. Andersen et al. [1].

The survival function is obtained from (1) as

$$S(t) = e^{-\int_0^t h(u) du}.$$

3.1.3 Components and repair

Some stochastic components in the engine are repaired when they fail to comply with the set of fitness rules. Some components are repaired at the military hangar and some are transferred to the repairbay at VAC. In the literature there are two extreme cases of repair, repair to a state *as good as new* or *as bad as old*.

The repair of a component to a state *as good as new* is often modeled with a renewal process. The idea of a renewal process is that we have a sequence of operating intervals of the length T_1, T_2, \dots . Each operating time ends with a failure and starts with a repair. If we have a stationary renewal process, each T_i has the same distribution. An alternative is to model with a non-stationary renewal process (NSRP). Then the T_i :s are modeled with different distributions, cf. Höyland and Rausand [17] and Nachlas [24]

The *as bad as old* or minimal repair is a repair process that repairs the component to the condition the component had just before the failure and is modeled with a non-homogeneous Poisson process (NHPP). Assume that N is a counting process that counts the number of repairs of a component and let the intensity (or hazard) function w be a function of time. If the process has independent and Poisson distributed increments, N is called a Poisson Process. If w is non-constant the process is non-homogeneous. If the number of repairs is Poisson distributed then N is called a non-homogeneous Poisson Process. For examples of NHPP with applications, cf. Höyland and Rausand [17], Nachlas [24] and Barlow and Hunter [5].

Imperfect repair models are models that describe component behavior somewhere between renewal and simple continuation. For examples of imperfect repair models, see Pham and Wang [26], Brown and Proschan [9], Bhattacharjee and Manish [7].

3.2 State data

The stochastic components in the aircraft engine are considered broken when they fail to comply with a set of fitness rules, e.g. limit of a crack lengths but also limits of damage made by corrosion or number of pieces burnt away. A component can be considered broken when any of those rules are violated. This means that the same type of components can fail in different ways. To make an observation of the state of the component is to observe the properties that make the component fail or possibly something that is related to those properties. When the properties are observed an estimate of the survival of the component can be obtained by using a model that predicts how the property is changing over time. It is possible to divide such models into two classes, empirical and mechanistic models. Empirical models aim to model the phenomena without using physical models while mechanistic models aim to model the physics behind the phenomena. All models are approximations, and this applies to mechanistic models as well as to empirical models.

In paper C an empirical model is constructed. The reason for choosing an empirical model instead of a mechanistic model is that the environment is too complicated

to model.

3.2.1 Empirical models

Empirical models are based on observations. When you one uses empirical model one believes that the future will behave the same way as the past. The strength of empirical models is that they can model and make predictions in an environment that is hard to model using known physics. Phenomena in reality are often stochastic in nature and with an empirical model it is possible to catch this randomness in a simple way by introducing stochastic variables. Empirical models are essential to use when the phenomena that we want to model are so complicated that known physical laws cannot model them or the modeling gets to complicated. The downside of empirical models is that since they do not care about the physics, they can seldom be moved to similar problems. Another downside is that they are based on observations and can hence not be used in an early phase where no observations are available, example the design phase of a component. It is also dangerous to extrapolate the models and use them in areas where they have not been supported by observations.

3.2.2 Mechanistic models

Mechanistic models are used when the phenomena that we attempt to model are well known, often in a more general perspective than the case we are trying to model. Mechanistic models have the advantage that they can be moved to other problems since they tend to be based on physical properties instead of observations. They can also be used in a very early planning phase of a product when no observations are made. Mechanistic models require knowledge about the environment boundary conditions.

3.3 Illustration of life estimation with state and survival data

For a specific component, state data gives considerably more information about the time when the component is going to fail than knowledge about survival data alone. We illustrate how the model in paper C can be used to improve the precision of the input distribution of the optimization model in paper D. Assume we are in a situation where we want to estimate time to failure for a component. We want to predict when the crack of the component reaches a length of 30 mm. At this crack length the component is considered to be so damaged that it has to be replaced. We want to make predictions at different times and with different amount of state data and we want to compare the predictions we get by using survival data alone. Assume that a large amount of survival data and a large amount of state data have led us to believe that the density function for the survival for new components is the one described in Figure 2. For new components the density is the same for both data sets. We now update this distribution with the new information we get. In

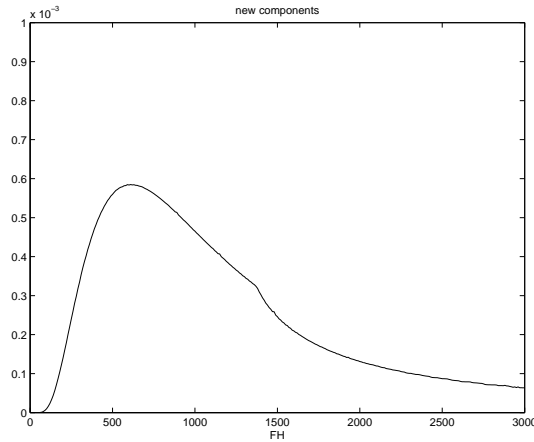


Figure 2: *Distribution of the time to failure for new components.*

the first case, the survival data case, we only know if the component is still alive. We will thus update the distribution with that information. In the second case we observe the crack length of the component and we will use that information to update the distribution. The observations of the crack are, $[0, 20, 25, 27, 33]$ mm at time $[200, 400, 600, 800, 1000]$ FH respectively. From the crack observations we understand that the crack growth speed is decreasing with time and that the crack passed the length of 30mm between 800 and 1000 FH. Hence this specific component should be considered broken somewhere between 800 and 1000FH.

Figure 3 illustrates the distribution of remaining life when the information known at 200 FH (to the left) and 400FH (to the right) has been processed. The upper left part of the picture illustrates the remaining life when only the fact that component is alive at 200 FH has been considered. In the lower left part the information that the crack length is 0 has been used, as the component is young and there is still no indication of a crack and the distributions are similar. At 400 FH the crack has started to grow and it is observed to be 20 mm. The upper right picture does not take the knowledge of the arising crack into account. It only acts as a model based on survival data would. The picture below is updated with the new information about the crack length and the estimated survival is lowered considerably.

Figure 4 illustrates the distributions after 600 FH (to the left) and 800 FH (to the right). More information about the crack growth is now known and the estimation of the time when the crack is going to reach 30 mm is even more precise. At 800 FH the crack observation says that the crack is almost 30 mm, and updating the estimation with that information gives a great advantage in comparison to not using it. The variation of the estimated distribution when the crack reaches 30 mm is large

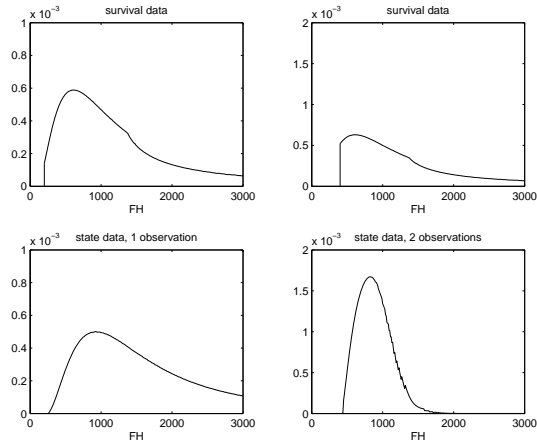


Figure 3: *Distribution of the time to failure at 200 FH to the right and 400 FH to the left. The distributions in the upper pictures have been updated with only the information of that the component is alive and the distributions in the lower pictures are based on crack observations and a crack growth model.*

even if we have a large number of observations. This is the combined result of the fact that the crack growth speed is decreasing and the fact that the observations are associated with an observation error.

4 Maintenance optimization

There exist numerous maintenance models in the literature, for a review see Barlow and Prochan [4], Pierskalla and Voelker [27], Sherif and Smith [30], Cho and Parlar [10], Frenk Dekker Kleijn [15], Aven and Dekker [3], Deeker and Wildman [12], Moraru and Popescu [23] and Andréasson [2].

The optimization model studied in paper A and used in a stochastic version in paper D is meant to be used for opportunistic maintenance. Opportunistic maintenance refers to the situation where preventive maintenance is carried out at opportunities.

In the problem at hand we have an engine system consisting of two kinds of components, deterministic components with a fixed life and stochastic components which fail randomly. The goal is to minimize the cost of maintaining the engine during a fixed predetermined time period. The fixed period may correspond to the planned life of the engine or the period where the company is responsible for maintaining the engine. There is a cost for replacing a component with a new one that corresponds to the value of the component and the work carried out to replace it. There is also

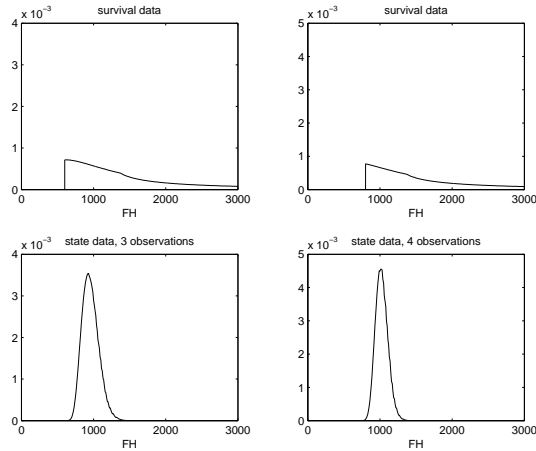


Figure 4: *Distribution of the time to failure at 600 FH to the right and 800 FH to the left. The distributions in the upper pictures have been updated with only the information of that the component is alive and the distributions in the lower pictures are based on crack observations and a crack growth model.*

a cost triggered for replacing any component at all, related to the cost of having the engine out of service and transport it to the repair bay and initiate repairs.

When a deterministic component has reached its predetermined life or a stochastic component has failed, the engine is forced to undergo maintenance at the repair bay. At this time a decision of what components should be replaced has to be made. There is an opportunity to do preventive maintenance and replace more than the failed component and hence increase the time to the next repair. What an optimal solution is depends on the relative costs of the components and the cost of bringing the engine to service.

The optimization problem above is stochastic because the life of the stochastic components are random and can be formulated

$$z = \min_{x \in X} \int f(x, u) dG(u) = \min_{x \in X} E[f(x, U)] \quad (7)$$

where $f(x, u)$ is the cost when using action x and the random outcome is u . We want to find the minimal expected cost z when G is the distribution of U and X some restriction of x .

It is possible to solve this type of problem with dynamic programming, see Bellman [6], but when the number of possible system states increases, dynamic programming becomes unsatisfactory to use. The problem can then be represented as a 0-1 linear programming problem cf. Andréasson [2].

4.1 Discretization and scenarios

In the formulation of the optimization model, the time span is divided by several equidistant points. The discretization impacts how the stochastic components can be represented in the optimization since the continuous life distribution has to be replaced by a discrete version.

When a suitable discretization has been found on the form (p_i, u_i) where u_i is the point of support with mass p_i , problem (7) reduces to

$$z = \min_{x \in X} \sum f(x, u_i) p_i. \quad (8)$$

Even with this representation there is a need to decrease calculation time by decreasing both the time resolution, that is longer distance between the equidistantly points, and also decrease the number of points of support in the representation of the life distribution. Furthermore, as the point of support has to coincide with the equidistant points this constrains the discretization even more. The issue of finding a discretization that suits the optimization model is addressed in paper D.

Finding a discretization of U and model with it as in (8) is called to create scenarios. Each outcome u_i is a possible future scenario. Creating a scenario tree for a component that covers the entire planned life of the engine involves not only discretization of the first stochastic component but also the stochastic component that replaces the first and so on. In paper D only discretization of the first stochastic component was considered. The replacing stochastic component was modeled as deterministic with $u_1 = E[U]$ and $p_1 = 1$.

For further discussion of stochastic optimization with scenarios, see Kall and Wallace [18].

4.2 Multistage problems

Modeling over time where decisions have to be made on several occasions makes it natural to introduce stages. Assume we have to make decisions today that influence the decisions tomorrow that affect the future. We can then introduce a multi-stage model with three stages: today = 0, tomorrow = 1 and future = 2. Let f_0 denote our cost function today and let $f_{1|x_0}$ denote the cost function tomorrow, which depends on x_0 , the decision today. The cost function in the future $f_{2|x_0, x_1}$ depends on both the decisions we make today and tomorrow. In this case the three-stage problem can be written.

$$z = \min_{x_0 \in X_0} f_0(x_0) \quad (9)$$

where

$$f_0(x_0) = \min_{x_1 \in X_1(x_0)} f_{1|x_0}(x_1) \quad (10)$$

and

$$f_{1|x_0}(x_1) = \min_{x_2 \in X_2(x_0, x_1)} f_{2|x_0, x_1}(x_2) \quad (11)$$

where x_1, x_2, x_3 are variables and X_1, X_2, X_3 are their domains.

The optimization model used in paper D is formulated as a two-stage problem which has integer restrictions on the x variables. A problem with integer restrictions is much harder to solve than the problem arising if the integer restrictions are removed. For further discussion of stochastic optimization with stages, see Kall and Wallace [18].

5 Summary of Paper A

In the aircraft industry maximizing availability is essential. Maintenance schedules must therefore be opportunistic, incorporating preventive maintenance activities within the scheduled as well as the unplanned ones. At the same time, the maintenance contractor should utilize opportunistic maintenance to enable the minimization of the total expected cost to have a functional aircraft engine and thus to provide attractive service contracts.

This paper provides an opportunistic maintenance optimization model which has been constructed and tested together with Volvo Aero Corporation in Trollhättan, Sweden, for the maintenance of the RM12 engine. The model incorporates components with deterministic as well as with stochastic lives.

The replacement model is shown to have favorable properties; in particular, when the maintenance occasions are fixed the remaining problem has the integrality property, the replacement polytope corresponding to the convex hull of feasible solutions is full-dimensional, and all the necessary constraints for its definition are facet-inducing.

Assuming that fatigue crack is the underlying failure mechanism, the model from Paper C that calculates the distribution of remaining life using an empirical crack growth model, is presented. Furthermore the empirical study in Paper B is presented and indicates that a non-stationary renewal process with Weibull distributed lives is a good model for the recurring maintenance occasions.

Assuming that fatigue crack is the underlying failure mechanism, and using the empirical crack model in Paper C is presented. Furthermore the empirical study in Paper B is presented and indicates that a non-stationary renewal process with Weibull distributed lives is a good model for the recurring maintenance occasions.

Using only one point of support for the distribution yields a deterministic replacement model; it is evaluated against classic maintenance policies from the literature through stochastic simulations. The deterministic model provides maintenance schedules over a finite time period that induce fewer maintenance occasions as well as fewer components replaced compared to the classic policies.

The error measure from paper D is presented and a study with several points of support showing that even more can be gained by using more than one point of

support.

6 Summary of Paper B

Military aircraft engines can offer greater operational availability and be more economically maintained through the use of better models to predict times to failure. Two models are used to analyze data gathered from Volvo Aero Corporation in Trollhättan. We are interested in the failure time distribution of the flame holder in the new RM12 engine. We have limited knowledge about it due to the limited number of RM12 engines currently in service. We have a large data set containing repair and maintenance times for the same type of details in the older RM8 engine. This paper will not discuss how to transform knowledge of the RM8 engine to the RM12 engine but will instead predict repair or maintenance times for the RM8 engine.

The first model is a non-stationary renewal process (NSRP) and the second is a non-homogeneous Poisson process (NHPP). We are interested in estimating the survival function and the hazard rate. In the NSRP we make a non-parametric estimation of the survival function with the help of the Kaplan-Meier estimator, cf. Høyland and Rausand [17]. We use kernel smoothing, cf. Klein and Moeschberger [19], to make a visual illustration of the density function. When the density function has been estimated, a parametric model is chosen to describe the times between repairs. Different estimation methods are discussed. In the NHPP we use the Nelson Aalen estimator to estimate the cumulative hazard function, cf. Andersen et.al. [1]. The NHPP is using the minimal repair assumption, cf. Høyland and Rausand [17], while in the NSRP the time to first repair is independent of the time to second repair.

An error measure is defined to compare how well the different models are suited to model current data, and we conclude that the NSRP process is the better model.

The NSRP process is developed to model different kinds of repair stations. There is one repair station at VAC and smaller repair stations closer to the hangars. With data we show that if engines are repaired at the smaller repair stations the time to next repair is generally shorter compared to if the engines are repaired at the VAC station. Finally we make two tests to see if the components are aging, but no aging is present other than a substantial degeneration after the first repair.

7 Summary of Paper C

In laboratory studies the stress strain cycles and other similar parameters are often considered known. In real word applications this is not always the case. Still we want to predict the remaining life of components.

In a case study of the crack growth of the low pressure turbine nozzle is studied. The historical data we have is limited and we do not know the exact nature of the crack growth. The temperature cycles in the engine result in the nozzle experiencing

fatigue and ultimately cracking. A nozzle is considered broken when the largest crack is above a certain level described in a set of fitness rules.

If we know what missions the plane is going to experience in the future, we can use those mission profiles and a thermodynamic model to try to estimate the loads. In this case we are not sure about future missions and we do not have a complete understanding of the thermodynamic environment. When a crack starts growing we get airflow through the crack that complicates the heat profile even more. Moreover, the property of the material changes through oxidation in the crack further complicating the physics behind the crack growth. Even without the difficulties above crack growth is stochastic in nature, c.f Virkler et al. [32], Bolotin [8], Yang [34]. If we want to predict the crack growth and control uncertainties we need a stochastic model.

Observations of the nozzle cracks show that the crack growth rate decreases as the crack length increases. D. B. Garcia et al. [16] study decreasing crack growth in an aircraft nose landing gear drag brace fitting by using a mechanistic model.

In order to model the crack growth and make predictions of the distribution of the remaining time to failure, a new empirical crack model is constructed. The variation of the basic components of the model is described by two stochastic variables describing the initiation time of the crack and the crack growth rate. A profile likelihood approach is used to determine both the distribution of the time when the crack reaches a certain size and the distribution of the crack size at a certain time. Related work using a profile likelihood approach has also been used in Lorén and Lundström [21].

8 Summary of Paper D

Aircraft engines can be more economically maintained and resources can be saved if the maintenance process is optimal. It can be a hard decision to decide what components in an engine to replace when the engine is being maintained. Several optimization models have been developed to deal with this problem. Epstein and Wilamowsky [13] and Dickman et al. [14] have developed models for modeling components with predetermined deterministic lives. Andréasson [2] has developed a model for details with deterministic lives as well as details with stochastic lives. Another approach to the problem is to construct a maintenance policy that is not always optimal but hopefully good. A survey of replacement and maintenance policies can be found in Wang [33].

We use the model presented in Andréasson [2] and formulate it as a two-step model. In order to use stochastic components in the optimization, the density functions need to be in a discrete form and there are also restrictions on what points of supports are allowed. An error measure closely related to the model is formulated. The error measure quantifies the cost of using different kinds of discretizations and the size of the error is related to the sup-distance between the distribution and

discretization. Four discretization methods will be presented and adapted to the constraints. The first method is the bracket mean method, cf. Smith [31], the second method minimizes the Wasserstein distance, cf. Pflug [25] and the third method keeps the moment of the distribution, cf. Miller and Rice [22]. The last method is a method that minimizes the sup-distance.

In order to keep down calculation times we want as few points of support as possible. Test runs with a Weibull distribution are done to compare the different discretization methods and conclude how the choice of number of points of support affects the accuracy. In the test the method that minimizes the Wasserstein and Sup-distance performed best. The method that preserved the moments performed worse. The error decreases with the number of support points. The minimum number of points of support that is suggested is three.

Finally the consequence of using a narrow scenario tree is discussed. A component that is exponentially distributed is sometimes replaced although the component is not aging.

9 Improvement of data gathering

In order to make better predictions and more accurate models, improvements in the data gathering and storing can be made.

In the current work description at VAC, the state of the stochastic components is observed if the engine is dismantled in the repairbay. However, the observations are not stored unless the observations state that the component has broken a fitness rule. The state data used in this thesis was an exception and comes from a few pri-engines. A pri-engine is an engine that is used to a great extent to catch critical errors before the main fleet reaches the same usage level. Conditions of pri-engines are stored more carefully. If a system for handling and storing state data information was created it would be possible to make models predicting remaining life for more components.

Observations of the state of the components are also to some extent made at the military hangar. If this information was to become available it would further increase possibility to make models.

The survival data for covariates consists of accumulated numbers of events that a detail has experienced from creation to death or censoring, no matter if the detail has been repaired several times. Better analysis could be performed if we knew the accumulated value of the parameter at every repair.

Storing more than the accumulated value of the parameters would facilitate the use of more advanced models, cf. e.g. Roemer and Ghiocel [28] and Roemer and Kacprzyński [29]. The advantage of the extension to those ideas is that the information collected during flight can directly predict the remaining life. Possible problems with such models is the high level of noise and variation in data, cf. Krok and Ashby [20].

10 Future work

In order to get the optimization model to work with the stochastic components from RM12 more work needs to be done. The following are some suggestions for future research.

The lives of the stochastic components may be dependent among themselves but also affected by the age of deterministic components and the age of the engine itself. It is not obvious how to get this information from the existing data. There is also dependence between some parts because they were observed when the engines were sent in for repairing other failures.

If and when new data becomes available, new empirical or mechanistic models can be created for the remaining stochastic components. The accuracy of the current empirical model for the low pressure turbine nozzle may be increased by using a more mechanistic approach. Such an approach would need to evolve a combined stress, heat and fluid model in addition to a statistical model.

The question of how to handle several competing failure nodes on one component has not been addressed in this thesis. When more data is available it is possible that some components will experience different kinds of failures. Competing risk models would be able to model this behavior. Competing risk models deal with the concept that a component is subject to failure as a result of the action of failure processes, which are competing to be the cause of the failure.

In the first years of the project the calculation to solve the optimization problem was time-consuming. Recent research in finding faster algorithms and new software has made solving more complex problems possible but it is still problematic to solve the optimization problem with stochastic details. A first step is to be able to model stochastic details with several points of support in a larger problem than the problem in paper D. A second step would be to model more than the first failure of a stochastic component with several points of support. A third step would be to reformulate the model from a second stage model to a model with more stages.

When modeling several stochastic components with several points of support, questions would arise how the joint distribution should be represented in the optimization model. There would also be questions about what the scenario tree should look like.

References

- [1] P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993.
- [2] N. Andréasson. Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems. Licentiate thesis, Chalmers University of Technology, 2004.

- [3] T. Aven and R. Dekker. A useful framework for optimal replacement models. *Reliability Engineering and System Safety*, 58:61–67, 1997.
- [4] R. Barlow and F. Proschan. *Mathematical Theory of Reliability*. John Wiley & Sons, New York, NY, 1965.
- [5] Richard Barlow and Larry Hunter. Optimum preventive maintenance policies. *Operations Res.*, 8:90–100, 1960.
- [6] R. Bellman. *Dynamic programming*. Princeton University Press, 1957.
- [7] Manish C. Bhattacharjee. New results for the Brown-Proschan model of imperfect repair. *J. Statist. Plann. Inference*, 16(3):305–316, 1987.
- [8] V.V. Bolotin. Reliability against fatigue fractures in the presence of set of cracks. *Engineering Fracture Mechanics*, 53:753–759, 1996.
- [9] Mark Brown and Frank Proschan. Imperfect repair. *J. Appl. Probab.*, 20(4):851–859, 1983.
- [10] D. I. Cho and M. Parlar. A survey of maintenance models for multi-unit systems. *European Journal of Operational Research*, 51:1–23, 1991.
- [11] D. R. Cox. Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.
- [12] R. Dekker, R. E. Wildeman, and F. A. van der Duyn-Schouten. A review of multi-component maintenance models with economic dependence. *Mathematical Methods of Operations Research*, 45:411–435, 1997.
- [13] Y. Epstein, S. Wilamowsky. Opportunistic replacement in a deterministic environment. *Comput. and Opns. Res.*, 12:311–322, 1985.
- [14] B. Dickman et al. Modeling deterministic opportunistic replacement as an integer programming problem. *Amer. J.Math. and Mgmt.*, 11, 1991.
- [15] J. B. G. Frenk, R. Dekker, and M. J. Kleijn. A unified treatment of single component replacement models. Technical report, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1996.
- [16] D.B. W. Cutshall Garcia, J.W Cardinal, C.B Fohn, O.H Burnside, and J. Pendleton. Application of fracture mechanics and damage tolerance techniques to a nose landing gear drag brace fitting. *Presented at the International Conference: Residual Fatigue Life and Life Extension of In-Service Structures.*, 11th International Spring Meeting(June 2006), Paris.

- [17] A. Høyland and R. Rausand. *System Reliability Theory*. John Wiley & Sons Inc., 1994.
- [18] Peter Kall and Stein W. Wallace. *Stochastic programming*. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons Ltd., Chichester, 1994.
- [19] J. Klein and M. Moeschberger. *Survival Analysis*. Springer-Verlag Inc., New York, 1997.
- [20] M. J. Krok and M. J. Ashby. Condition-based diagnostic gas path reasoning for gas turbine engines. In *Proceeding of the 2002 IEEE International Conference on Control Applications*, pages 227–234. 2002.
- [21] S. Loren and M. Lundström. Modelling curved s-n curves. *Fatigue & Fracture of Engineering Materials and Structures*, 28(5):437–443, 2005.
- [22] T.R. Miller, A.C. Rice. Discrete approximations to probability distributions. *Management Science*, 29:352–362, 1983.
- [23] S. Moraru, F. Sisak, and I. Popescu. Overview of preventive maintenance policies for systems, generally, and for digital systems, particularly. *Proceedings of the 6th International Conference on Optimization of Electrical and Electronic Equipments*, 3:801–806, 1998.
- [24] Joel A. Nachlas. *Reliability Engineering, Probabilistic Models and Maintenance Methods*. CRC press, 2005.
- [25] G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program.*, 89(2, Ser. B):251–271, 2001. Mathematical programming and finance.
- [26] H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94:425–438, 1996.
- [27] W. P. Pierskalla and J. A. Voelker. A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly*, 23:353–388, 1979.
- [28] M.J. Roemer and D.M. Ghiocel. Probabilistic approach to the diagnosis of gas turbine engines. Technical report, 53rd Machinery Prevention Technologies (MFPT), Conference, Virginia Beach, VA, April 1999.
- [29] M.J. Roemer and G.J. Kacprzynski. Advanced diagnostics and prognostics for gas turbine engine risk assessment. Technical report, IGTI/ASME Turbo Expo, Munich, Germany, May 2000.

- [30] Y. Sherif and M. Smith. Optimal maintenance models for systems subject to failure - a review. *Naval Research Logistics Quarterly*, 28:47–74, 1981.
- [31] J.E. Smith. Moment methods for decision analysis. *Management Science*, 39:340–358, 1993.
- [32] D.C. Virkler, B.M Hillberry, and P.K Goel. The statistical nature of fatigue crack propagation. *Journal of Engeneering Materials and Technology*, 101:148–153, 1979.
- [33] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139:469–489, 2002.
- [34] J.N. Yang. A simple second order approximation for stochastic crack growth analysis. *Engineering Fracture Mechanics*, 53:677–686, 1996.

Paper A

Optimization of opportunistic replacement activities: A case study in the aircraft industry

Torgny Almgren,^{*} Niclas Andréasson,[†] Dragi Anevski,[‡]
Michael Patriksson,[‡] Ann-Brith Strömberg,[§]
and Johan Svensson[‡]

Abstract

In the aircraft industry maximizing availability is essential. Maintenance schedules must therefore be opportunistic, incorporating preventive maintenance activities within the scheduled as well as the unplanned ones. At the same time, the maintenance contractor should utilize opportunistic maintenance to enable the minimization of the total expected cost to have a functional aircraft engine and thus to provide attractive service contracts. This paper provides an opportunistic maintenance optimization model which has been constructed and tested together with Volvo Aero Corporation in Trollhättan, Sweden for the maintenance of the RM12 engine. The model incorporates components with deterministic as well as with stochastic lives. The replacement model is shown to have favourable properties; in particular, when the maintenance occasions are fixed the remaining problem has the integrality property, the replacement polytope corresponding to the convex hull of feasible solutions is full-dimensional, and all the necessary constraints for its definition are facet-inducing. We present an empirical crack growth model that estimates the remaining life and also a case studied that indicates that a non-stationary renewal process with Weibull distributed lives is a good model for the recurring maintenance occasions. Using only one point of support for the distribution yields a deterministic replacement model; it is evaluated against classic maintenance policies from the literature

^{*}Volvo Aero Corporation, SE-461 81, Trollhättan, Sweden.

[†]Tegnérgatan 29, SE-333 32 Smålandsstenar, Sweden.

[‡]Mathematical Sciences, Chalmers University of Technology, and Mathematical Sciences, Göteborg University, SE-412 96 Gothenburg, Sweden

[§]Fraunhofer-Chalmers Research Centre for Industrial Mathematics, Chalmers Science Park, SE-412 88 Gothenburg, Sweden.

through stochastic simulations. The deterministic model provides maintenance schedules over a finite time period that induce fewer maintenance occasions as well as fewer components replaced.

Contents

1	Maintenance activities at Volvo Aero Corporation	5
1.1	Preliminaries	5
1.2	The structure of the RM12 engine	6
1.3	Components and maintenance schedules	6
1.4	Contracts	8
1.5	Maintenance principles	8
1.6	Scope and outline	10
2	A deterministic, opportunistic maintenance model for an engine module	12
2.1	The model	12
2.2	Numerical illustration	13
3	The replacement polytope	15
3.1	Polyhedral combinatorics	15
3.2	The dimension and basic facets of $\text{conv } S$	18
3.3	A new class of facets: An example	20
3.4	Conclusions	21
4	Towards a model for the stochastic optimization problem	22
4.1	Structural models for time between repairs	24

4.1.1	Survival analysis	25
4.1.2	Model fit	26
4.1.3	Parametric modeling and likelihood based inference	30
4.1.4	The one dimensional case: Crack length modeling	31
4.1.5	Case study	33
4.1.6	Model illustration	33
4.2	An optimal discretization of a continuous distribution	36
4.2.1	Error measure	39
4.2.2	Discretization approaches	40
4.2.3	Test results	43
4.3	Output and relation to the optimization problem	44
5	Maintenance policies	46
5.1	A value policy	47
5.2	An age replacement policy	47
6	Simulations	48
6.1	The deterministic problem	49
6.2	Stochastic simulations	50
7	Conclusions	53
A	Proofs of the main results of Section 3	56

Introduction

Industrial activities are often characterized by the use of very expensive equipment that needs to be utilized as efficiently as possible to pay back the cost of investment. This essentially means that the equipment should be used with as few and short interruptions as possible. Typical examples are power plants (e.g., nuclear plants), processing industry (e.g., paper plants) and the aviation industry. A vital part of the latter case is concerned with the maintenance of aircraft engines.

When an aircraft engine is removed for overhaul, it needs to be replaced by a spare engine to facilitate the use of the airframe as it is the operator's main interest to have access to operational aircrafts during the maintenance period. This is normally achieved by the use of spare engines. These engines could be owned by the operator or the maintenance supplier, but also be leased from a third party. The cost for the spare engine is always high, irrespective of how it is obtained. Every maintenance event is therefore associated with a large, more or less fixed, cost in addition to the variable cost (e.g., material costs). As this fixed cost is independent of the actions that are performed, there is a need to consider that the maintenance event is an opportunity for preventive maintenance—an opportunity that should be used in an optimum way! In essence the cost for production interruption must be balanced versus the variable cost of the maintenance event. (This is often denoted opportunistic maintenance, cf. [11].)

An aircraft engine consists of thousands of parts. Some of the parts are safety-critical, which means that if they fail there will be an engine breakdown, possibly with catastrophic consequences. Therefore, the safety-critical parts have fixed life limits, and must be replaced before these are reached. Hence we consider the safety-critical parts as having deterministic life limits. These limits are measured in "cycles" and are strictly regulated. All other parts of the engine are considered to have stochastic lives. The problem with them is that their lives need to be estimated, which makes it difficult to compute a reliable replacement schedule. For some of these parts failure distributions may be computed from historical data and monitoring observations. This information could then be discretized and be used as an input into optimization models. Section 4 contains a detailed discussion on this topic.

When a deterministic life limit is reached, or when there is another indication that the engine is not performing as it should, the engine must normally be taken out of service and sent to the workshop. This is, as earlier indicated, an opportunity for preventive replacements of non-failed parts with stochastic lives and of deterministic parts that have not yet reached their respective life limits! An issue at this point is thus to know which actions should be taken and which parts should be replaced.

A current trend in service workshops in the aircraft industry is to offer the complete undertaking of the maintenance of all engines belonging to the customer. This results in contracts where the customer pays a fixed price per flight hour and the maintenance supplier ensures access to a working fleet of engines throughout the contract period. The ability to offer attractive contracts is therefore to a large extent dependent on the actual flight hour cost that can be achieved by the use of good planning practices. When the maintenance contract has been signed, the profit for the supplier obviously is directly related to how well the maintenance is carried out.

When the time period for a maintenance contract runs out it is typically advantageous for the workshop that the remaining lives of the parts of the engine are small (at least if a sequel contract has not been signed). Contracts however often describe how the difference in engine status, between start and end of contract period, should be regulated. How the value of this status should be computed must therefore be stated within the contract, so that it can be taken into consideration when the maintenance is planned.

In this article we develop maintenance optimization models to minimize the total expected cost to have a functional aircraft engine (consisting of parts with deterministic life limits and stochastic lives) during a finite time period (such as the contract period or the expected life span of the engine). The output from these models is replacement schedules for each maintenance occasion. The optimization models are however primarily intended to be used to determine a preliminary work scope when the aircraft engine is taken to the service workshop.

1 Maintenance activities at Volvo Aero Corporation

1.1 Preliminaries

VAC (Volvo Aero Corporation, Trollhättan, Sweden) manufacture and maintain the RM12 engine, which is the engine of the military aircraft JAS 39 Gripen. Gripen is mainly used by the Swedish Air Force (SAF), whose fleet encompasses about 200 RM12 engines. The discussion below is mainly restricted to the RM12 and the relationship between VAC and SAF where SAF and VAC jointly strives for as low total flight hour cost as possible.

The RM12 engine consists of several modules, each comprising several components (the modular concept is briefly discussed in [14]). The modules, that each contain a number of components or parts, can individually be removed (and replaced), and shipped to and from the workshop. When a component is to be replaced the corre-

sponding module is, if required, sent to the service workshop.

Some of the parts in the RM12 engine are life limited. The life limits of these parts are measured in the number of "cycles" they may be used. For a given part this number depends on the load profile during the use of the engine up to that time point, so when the engine is driven hard the number of cycles accumulates faster. The life limits are calculated such that the probability that a part fails before its estimated life limit is over is lower than one per mille.

1.2 The structure of the RM12 engine

In order to remove a specific part from a module it is most often necessary to remove other parts as well. Figure 1 illustrates the structure of the deterministic parts of the RM12 engine. Often there are several ways to reach a specific part. According

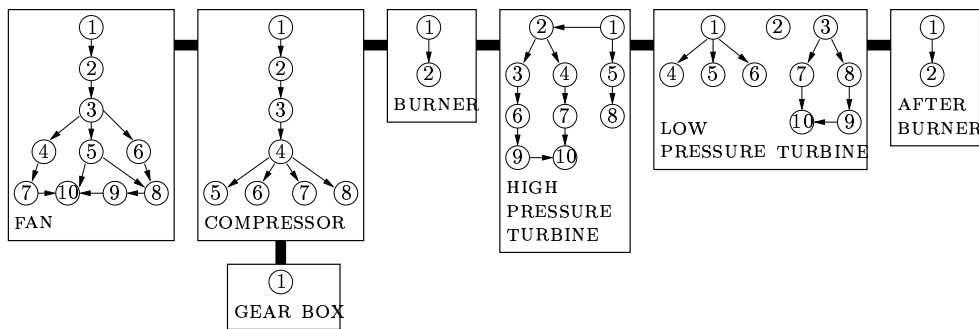


Figure 1: A graphical representation of the deterministic parts of the RM12 engine. Each box represents a module, each node part, and each arc a possible path to reach the part towards which the arc is pointing.

to Figure 1 there are two possible ways to remove part 8 in the fan module. First, one has to remove parts 1, 2, and 3 (in this order), then either part 5 or part 6, and finally part 8.

1.3 Components and maintenance schedules

The maintenance of aircraft engines is either planned or un-planned (on condition). In each engine there are sensors at different locations that continuously measure, for example, pressure, temperature, the number of ignitions, and the number of cycles

accumulated for each part, which are also kept on record. This data is used to establish when on condition maintenance need to be performed, but also supplies the basis for the life usage calculations.

A need for maintenance (or replacement) appears when a part reaches its life limit, fails or if the engine monitoring system indicates that the engine does not perform as well as it should. Unplanned maintenance also occurs due to unexpected events as accidents—sometimes birds are sucked into the turbine and through the engine, causing heavy damages—or the failure of a part with stochastic life. When this happens SAF places a maintenance order at VAC. The engine or module(s) that needs to be serviced is then sent to the service workshop.

When a module arrives at the service workshop at VAC the preliminary work scope is determined. Inspection, using advanced techniques, such as fiber optics, can be used at this stage. The module is then disassembled to the level required; parts are removed, cleaned and further inspected. A decision of the final work scope (e.g., which components to replace etc) is then decided jointly by SAF and VAC.

Repair times, but especially the delivery times, for new replacement parts, are often very long. Because of this, components are often replaced by components from stock to save time. Both new and used components are kept in stock, where the used parts generally have a shorter remaining life span.

When the engine is taken to the workshop in order to replace any part there is an opportunity to replace also parts with stochastic lives that have not yet failed and deterministic parts that have not yet reached their estimated life limits. This is often denoted opportunistic maintenance ([11]) and is mainly motivated by the fixed cost— independent of which parts that are replaced—associated with taking the engine to the workshop.

When the engine is at the workshop the parts with stochastic lives are inspected and their respective conditions are estimated. Based on this estimation and historical data their failure distributions can be computed using methods described in Section 4. The optimization model computes what to replace at the specific maintenance occasion in order to minimize the total expected maintenance cost within the planning horizon, given inputs from the failure distributions and the remaining lives of the deterministic parts, as well as material costs of new parts, work-cost to replace parts, etcetera.

At every maintenance event a new optimization of the maintenance schedule is performed. Each time a part with a stochastic life is inspected more information is also received about its condition; its failure distribution can then be updated, which results in a smaller variance.

To summarize, the optimization model described in this paper aims at minimizing the total expected cost during a given time period. The developed optimization model is designed to consider the cost for interrupted production while minimizing the cost of maintenance. In practice meaning that the model will strive to create a maintenance plan with as infrequent maintenance occurrences as possible while maintaining a sound use of replacement parts, new as well as used parts.

1.4 Contracts

When the time period for a maintenance contract runs out it is typically advantageous for the workshop that the remaining lives of the parts of the engine are small (at least if a sequel contract has not been signed). The earlier a customer signs a new contract the better the maintenance activities can be planned. A reasonable policy for the workshop is to give the customer some type of discount if a new contract is signed before the current contract runs out.

An optimization model aims at minimizing the total expected cost during a given time period, so if this equals the contract period the model will tend to make use of the values of the parts as well as possible. The required status of the engine at the end of the contract period (which depends on the remaining lives of the parts) can be given in the contract. It can then be considered as a constraint in the optimization model (in fact, this is a type of availability constraint). It is also possible to assign a value to the engine (for the workshop) at the end of the contract period that depends on the remaining lives of the parts. How this value shall be computed must then be given in the contract, so that it can be taken into consideration when the maintenance is planned.

1.5 Maintenance principles

The literature on maintenance principles has been reviewed; see [3, Chapter 8]. We provide here a brief summary of our findings. (For general reviews, see [6], [27], [29], [9], [16], [5], [12], [23].)

Under an *age replacement policy* a component is replaced at failure or at a specified age, whichever occurs first. The basic age replacement policy is described in [6]. Fox [15] refines the age replacement policy by incorporating discounting, that is, the loss incurred at a replacement decreases with time. This model is further investigated by Ran and Rosenlund [28], who also perform a sensitivity analysis, and give some numerical examples. Age replacement policies can also be governed by condition monitoring devices; see, e.g., Kumar and Westberg [20].

Under a *block replacement policy* the components of a system are replaced at failure or at fixed times kT ($k = 1, 2, \dots$), whichever occurs first. The basic block replacement policy is described in [6]. The main drawback of the block replacement policy is that practically new items might be replaced at planned replacement times. Berg [7] presents a modified block replacement method where failed items are still replaced after failure, but items possessing the age $b \in [0, T)$ or less at a scheduled block replacement point are not replaced by new items. The objective is to find b and T such that the expected cost per unit time and item over an infinite time horizon is minimized. Archibald and Dekker [4] extend the modified block replacement policy presented in [7] in two ways. They consider (i) a discrete time framework which allows the use of any discrete life distribution, and (ii) multi-component systems. The case when the system consists of identical components is treated in detail, and an example with a Weibull life distribution is presented. The authors outline how to extend the model to multi-component systems with nonidentical components.

Sometimes failed components can be detected and replaced only by inspection. There is a cost related to the time a component is not operative. Under an *inspection policy* the objective is to find the inspection schedule that minimizes the expected average cost.

Maintenance policies such as age and block replacement are examples of *scheduled maintenance policies*. These policies are easy to implement since they have a clear structure. Nevertheless, often *condition based maintenance* can be better and more cost effective. Under a condition based maintenance policy a technical state of the system is monitored or inspected, and when a specific threshold value is reached the system is replaced or preventive maintenance is performed. The principles and implementations of condition-based preventive maintenance are discussed by Mann et al. [19], who also compare the age replacement and condition-based policies.

Opportunistic maintenance refers to the situation in which preventive maintenance is carried out at opportunities. In the literature it is sometimes assumed that these opportunities arise independently of the failure process; sometimes the opportunities are by definition equal to failure epochs of individual components. In the latter case, due to economies of scale (for example, fixed costs at each maintenance occasion independent of what is replaced), the unpleasant event of a failing component is at the same time considered as an opportunity for the preventive maintenance of other components. This situation is typical for the maintenance of aircraft engines.

Finally, we refer to previous application work on aircraft engine maintenance. (A deeper treatment of the theory and the applications presented below in [17] and [18] can be found in [21].)

Hopp and Kuo [17] study the *maintenance of an aircraft engine* by dividing the parts

into non-safety-critical and safety-critical parts. The non-safety-critical parts do not fail, but the cost of loss of performance increases with age. A safety-critical part has a life distribution, and when it fails it destroys the whole system but it has no associated cost for loss of performance. The authors draw the conclusion that optimal policies are likely to be extremely difficult to compute and—since their form is complex—very difficult to communicate and use in practice. Therefore, heuristics are suggested for the case of a system with zero or one safety-critical component and multiple non-safety-critical components. Lower bounds on costs are computed to evaluate the performance of the heuristics.

In [18] the *maintenance of the compressor* of an aircraft engine is considered. It is assumed that fatigue crack is the underlying failure mechanism and the crack growth is due to the number of “shocks” monitored by sensors. The available information about the crack growth process is the crack size observed at the most recent inspection/replacement and the number of shocks experienced since then. At the beginning of each flight it is decided—based on the observed state and the number of shocks to be incurred during the flight—whether or not to schedule an inspection at the end of the current flight. After inspection the true crack size will become known, and it must be decided whether a blade replacement is needed or not. A dynamic programming recursion for the problem is developed. The authors point out that a general policy from a complex dynamic program can be difficult to compute and communicate, and therefore it is useful to characterize the optimal policy as having some kind of simple structured form. This turns out to be possible for the compressor maintenance problem. (Crack growth modelling and monitoring is also a basis of our maintenance model.)

1.6 Scope and outline

The main part of this article deals with the development of optimization models for the maintenance of multi-component systems consisting of parts with deterministic or stochastic lives. In the main part of the related literature one assumes that the systems consist of parts with stochastic lives only, the time horizon is infinite, and a policy is used to find a replacement scheme. Also, it is clear from the literature that it is extremely hard to find an optimal replacement schedule when the number of parts is large, and hence different replacement policies are developed. Such policies reduce the complexity of the problems, but the solutions found are most often not optimal. Further, the literature points out that the case of a finite time horizon is even harder than the infinite time horizon case.

In our aircraft application the time horizon is finite and the number of parts is large, so if all of them were stochastic it would be necessary to use replacement

policies. However, about 75% of the components considered in an aircraft engine are deterministic, so our problem is more structured than the completely stochastic systems considered in the literature.

The contribution of this paper is three-fold. First, we provide a linear integer opportunistic maintenance model for aircraft engine modules, based on the replacement model presented in [13]; we establish its advantages over simpler policies from the literature and current practice at VAC in providing good schedules. Second, we establish attractive mathematical properties for its efficient solution; this is especially important because in a future development, maintenance schedules are to be optimized for entire engine fleets, wherein the model developed here will be a sub-model. Third, we establish statistically valid methodologies for incorporating parts with stochastic lives in our model, through the estimation of their remaining lives.

The remainder of the paper is organized as follows. In Section 2 we present a mathematical model for generating optimal replacement schedules over finite time horizons and provide a numerical example showing the influence of fixed costs for the maintenance of a module on the importance of opportunistic maintenance. In Section 3 we perform a polyhedral study of the convex hull of the set of feasible solutions to this model, referred to as the replacement polytope. We show that the replacement polytope is full-dimensional under general assumptions. Also, we show that if the variables associated with the fixed costs in the model are fixed to integers, then the polyhedron arising from the continuous relaxation of the variables associated with the replacement of the parts is integral. The inequality constraints in the original formulation are studied and we show that several of them are facet-defining. Further, we show that the inequalities in the original formulation are not sufficient to completely describe the replacement polytope. By using Chvátal–Gomory rounding we construct a new class of valid inequalities and show that these inequalities (in some cases) are facet-defining. In Section 4 we outline survival estimation models, and show how measurements of crack development in parts with stochastic lives can be used to define, and enrich, Weibull distributions for the estimation of conditional life distributions. Section 5 presents the current maintenance policy used at VAC, as well as an age replacement policy; Section 6 is devoted to a numerical study of the stochastic properties of the optimization model and of the above-mentioned policies in the form of stochastic simulations; it shows that the optimization model always is to prefer to simple policies, even when the uncertainty in the lives of the stochastic parts is quite substantial.

2 A deterministic, opportunistic maintenance model for an engine module

Consider a system consisting of N deterministic parts and a finite time horizon discretized into $T + 1$ time steps $t = 0, 1, \dots, T \geq 2$. At time step $t = 0$ all of the parts of the system are new and at time step $t = T$ the system will be discarded. We introduce the set $\mathcal{N} = \{1, \dots, N\}$. The life of a new part of type $i \in \mathcal{N}$ is $T_i \geq 1$ time steps and its purchase cost is $c_i > 0$ monetary units. There is a fixed cost of $d > 0$ monetary units associated with each replacement occasion, independent of the number of parts replaced at that occasion. The objective is to minimize the cost of having a working system between the time steps 0 and T .

2.1 The model

In order to formulate a linear integer programming model that solves the *replacement problem*, we introduce the variables

$$x_{it} = \begin{cases} 1, & \text{if part } i \text{ is to be replaced at time } t, \\ 0, & \text{otherwise,} \end{cases} \quad i \in \mathcal{N}, t = 1, \dots, T - 1,$$

$$z_t = \begin{cases} 1, & \text{if any of the parts } i \in \mathcal{N} \text{ is to be replaced at time } t, \\ 0, & \text{otherwise,} \end{cases} \quad t = 1, \dots, T - 1.$$

The variables x_{it} and z_t are not defined for $t \in \{0, T\}$, since it will never be beneficial to replace any part at these time points. To force the replacement of a part before its life limit is exceeded we define constraints considering the lives of the parts and fixed costs. Each part of the system has a fixed life limit and at the very latest, when this is reached the part must be replaced. A part $i \in \mathcal{N}$ with life limit T_i time steps, where $1 \leq T_i \leq T - 1$, must be replaced at least once every T_i time steps yielding the constraints

$$\sum_{i=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \dots, T - T_i, \quad i \in \mathcal{N}.$$

Every time the replacement of some part $i \in \mathcal{N}$ is triggered, a fixed cost must be paid, indicated by the variable z_t having the value 1, leading to the constraints

$$x_{it} \leq z_t, \quad i \in \mathcal{N}, \quad t = 1, \dots, T - 1. \quad (1)$$

REMARK 1 (strong formulation) The model presented in [13] includes the constraints

$$\sum_{i \in \mathcal{N}} x_{it} \leq N \cdot z_t, \quad t = 1, \dots, T-1, \quad (2)$$

instead of (1). However, as discussed in Section 3, the constraints (1) are stronger in the sense that the linear programming relaxation of the formulation including these constraints has a smaller feasible set than the one including the constraints (2). \square

If a part of type $i \in \mathcal{N}$ is replaced at time step t , the cost c_i must be paid. Further, if any of the parts $i \in \mathcal{N}$ is replaced at time step t , the fixed cost d must be paid. A complete model of the minimization of the total cost for having a working system between the time steps 0 and T is then given by

$$\text{minimize} \quad \sum_{t=1}^{T-1} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + dz_t \right), \quad (3a)$$

$$\text{subject to} \quad (x, z) \in S, \quad (3b)$$

where

$$S = \left\{ (x, z) \in \mathbb{B}^{N(T-1)} \times \mathbb{B}^{T-1} \mid \begin{array}{l} \sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \dots, T - T_i, \quad i \in \mathcal{N}; \\ x_{it} \leq z_t, \quad t = 1, \dots, T-1, \quad i \in \mathcal{N} \end{array} \right\}. \quad (4)$$

The model (3) is called the *replacement problem*.

2.2 Numerical illustration

We illustrate how the fixed cost d in the model (3) affects the structure of the optimal maintenance schedule.

Consider an instance of (3) with $T = 60$, $N = 4$, $T_1 = 13$, $T_2 = 19$, $T_3 = 34$, $T_4 = 18$, $c_1 = 80$, $c_2 = 185$, $c_3 = 160$, and $c_4 = 125$. The data is chosen so that the relations between the life limits and the costs are similar to those for the fan module of the RM12 engine. The model (3) is then solved for each of the fixed cost values $d = 0$, $d = 10$, and $d = 1000$ (in a real maintenance situation $d = 10$ is the most reasonable value among the three).

For $d = 0$, the total number of replacement occasions becomes 11. Since the fixed cost is zero there are no advantages with replacing components before their respective life limits are reached.

For $d = 10$, compared to the case $d = 0$ the total number of replacement occasions has decreases from 11 to five. It is now beneficial to replace the components in larger groups and they are often replaced before their respective life limits are reached.

Consider finally the case $d = 1000$. Since the fixed cost is high compared to the costs of the components themselves it is very important to utilize the opportunity to replace several components at the same time. The total number of replacement occasions is four. Actually, since $T_1 = 13$ and $T = 60$ there exists no feasible replacement schedule for which the total number of replacement occasions is less than four.

Figure 2 shows the maintenance occasions for the three cases. The horizontal axis represents the 60 time periods and each maintenance occasion is represented by a vertical bar, where a dot at a certain height represents a component of the corresponding type being replaced. The figure clearly illustrates how opportunistic maintenance becomes more beneficial with an increasing fixed cost.

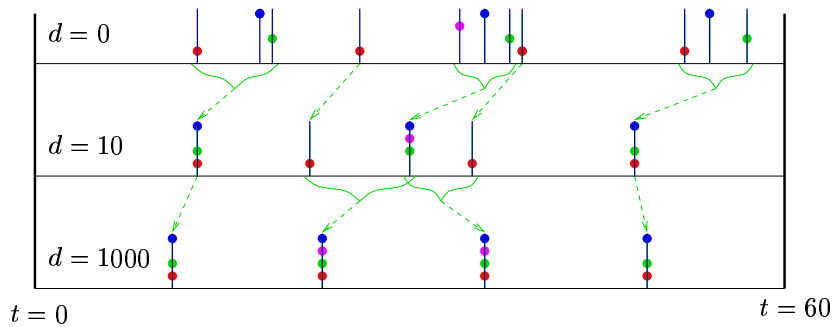


Figure 2: An illustration of the differences between optimal maintenance schedules for $d = 0$, $d = 10$, and $d = 1000$. When the fixed cost increases from 0 to 10 the first three replacement occasions for $d = 0$ are grouped into one for $d = 10$. The fourth replacement occasion for $d = 0$ is moved to an earlier time point, but still it is just component one that is replaced. Further, the replacement occasions 5–8 for $d = 0$ are grouped into two replacement occasions for $d = 10$. The three last replacement occasions for $d = 0$ are grouped into one for $d = 10$. Similarly, when the fixed cost is increased from $d = 10$ to $d = 1000$, the replacement occasions 2–4 for $d = 10$ are grouped into two replacement occasions for $d = 1000$.

3 The replacement polytope

In this section we study the structure of the set S , defined in (4), of feasible solutions to (3). The convex hull of S , denoted $\text{conv } S$, is called the *replacement polytope*. The main goal of studying the facial structure of S is to completely describe its convex hull by a finite set of linear inequalities which will make it possible to solve the problem using linear programming techniques. Our ambition here is to take the first steps towards such a complete linear description of the replacement polytope.

We first review some basic results on polyhedral combinatorics. Then we compute the dimension of the replacement polytope and conclude that some of the inequalities in the original formulation (4) define facets of the replacement polytope. However, using an example we show that these basic inequalities do not completely define $\text{conv } S$. We then derive a new class of facets by using Chvátal–Gomory rounding. We conclude the section with suggestions on further studies on the facial structure of the replacement polytope.

3.1 Polyhedral combinatorics

We here review the results on polyhedral combinatorics necessary for the derivation of our results on the facial structure of the replacement polytope. A comprehensive survey of polyhedral combinatorics is given in [24].

Let X be a subset of \mathbb{R}^n . The set X is an *affine set* if $\lambda x + \mu y \in X$ whenever $x, y \in X$ and $\lambda, \mu \in \mathbb{R}$ are such that $\lambda + \mu = 1$. A point $x \in \mathbb{R}^n$ is an *affine combination* of the points $x^1, \dots, x^m \in \mathbb{R}^n$ if there exist scalars $\lambda_1, \dots, \lambda_m$ with $\lambda_1 + \dots + \lambda_m = 1$ such that $x = \lambda_1 x^1 + \dots + \lambda_m x^m$. The *affine hull* of X , denoted by $\text{aff } X$, is the set of all (finite) affine combinations of points of X . The set X is *affinely dependent* if there exists an $x \in X$ such that $x \in \text{aff}(X \setminus \{x\})$. Finally, the dimension of the set X , denoted by $\dim X$, is one less than the maximum cardinality of an affinely independent set $K \subseteq X$.

A *polyhedron* in \mathbb{R}^n is a set of the form

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b\}, \quad (5)$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The *equality subsystem* $(A^=, b^=)$ of P is defined by the rows of the system $Ax \leq b$ that are fulfilled with equality for all $x \in P$. The matrix $A^=$ will be referred to as the matrix corresponding to the equality subsystem of P .

PROPOSITION 2 (dimension of a polyhedron, [24, p. 87]) *If $P \subseteq \mathbb{R}^n$ is a polyhedron, then*

$$\dim(P) + \text{rank}(A^=, b^=) = n.$$

□

If $\dim P = n$ we say that P is full-dimensional.

PROPOSITION 3 ([33, p. 81]) *Let V be a finite set in \mathbb{R}^n and let $X = \text{conv } V$. Then each extreme point of X lies in V .*

□

PROPOSITION 4 ([10, p. 206]) *Every polytope equals the convex hull of its extreme points.*

□

Another useful result is that every polytope is a polyhedron.

PROPOSITION 5 ([33, p. 114]) *A set is a polytope if and only if it is a bounded polyhedron.*

□

There is an obvious relation between the dimension of a set $X \subseteq \mathbb{R}^n$ and that of $\text{conv } X$.

PROPOSITION 6 *Let $X \subseteq \mathbb{R}^n$, then $\dim X = \dim(\text{conv } X)$.*

□

If all of the extreme points of a polyhedron are integral the polyhedron is called *integral*. A matrix is said to be *totally unimodular* (TU), if all of its square submatrices have the determinant 0, 1, or -1 . A sufficient condition for a polyhedron to be integral is given by the following.

PROPOSITION 7 ([10, p. 221]) *Let $A \in \mathbb{R}^{m \times n}$ be a totally unimodular matrix and let $b \in \mathbb{R}^m$ be integral. Then the polyhedron defined by $Ax \leq b$ is integral.*

□

We will utilize the following characterization of total unimodularity.

PROPOSITION 8 (characterization of the TU property, [24, pp. 542–543]) *Let A be a matrix in $\mathbb{Z}^{m \times n}$. The statements (i) and (ii) are equivalent:*

(i) A is TU ;

(ii) For every $J \subseteq \{1, \dots, n\}$ there exists a partition J_1, J_2 of J such that

$$\left| \sum_{s \in J_1} a_{rs} - \sum_{s \in J_2} a_{rs} \right| \leq 1, \quad r = 1, \dots, m. \quad (6)$$

□

Let P be given by (5). The inequality $\pi x \leq \pi_0$ is called a *valid inequality* for P if it is satisfied by all points in P . If $\pi x \leq \pi_0$ is a valid inequality for P , and

$$F = \{x \in P \mid \pi x = \pi_0\},$$

then F is called a *face* of P , and we say that $\pi x \leq \pi_0$ *defines* F . A face F of P is said to be *proper* if $F \notin \{\emptyset, P\}$. A face F of P is called a *facet* of P if $\dim F = \dim P - 1$. It holds (cf. [24, p. 89]) that if F is a facet of P , then there exists some affine inequality defining F .

We now arrive at the crucial result that every full-dimensional polyhedron can be uniquely represented by its facets.

PROPOSITION 9 ([24, p. 91]) *A full-dimensional polyhedron P has a unique (to within scalar multiplication) minimal representation by a finite set of linear inequalities. In particular, for each facet F_i of P there is an inequality $a^i x \leq b_i$ (unique within scalar multiplication) representing F_i and $P = \{x \in \mathbb{R}^n \mid a^i x \leq b_i, i = 1, \dots, k\}$. □*

Propositions 4 and 5 imply that if $X \subseteq \mathbb{R}^n$ is a finite set, then the polytope $\text{conv } X$ is a polyhedron. Hence, if $\text{conv } X$ is full-dimensional, from Proposition 9 it follows that the union of all facet-defining inequalities of $\text{conv } X$ defines a linear description of it. Therefore, it is of interest to find facets of a polytope defined by inequalities and integrality constraints. The following characterization, based on the uniqueness property in Proposition 9, is useful when proving that a certain valid inequality is a facet.

PROPOSITION 10 ([24, pp. 91–92]) *Let P be a full-dimensional polyhedron and let $F = \{x \in P \mid \pi x = \pi_0\}$ be a proper face of P . Then the following two statements are equivalent:*

(i) F is a facet of P ;

(ii) If $\lambda x = \lambda_0$ for all $x \in F$, then $(\lambda, \lambda_0) = \alpha(\pi, \pi_0)$ holds for some $\alpha \in \mathbb{R}$. \square

We close this section by remarking that it follows from Proposition 3 that all of the extreme points of $\text{conv } S$ belong to S . Hence, if we can find a polyhedral description of $\text{conv } S$, then the replacement problem (3) can be solved by standard linear programming techniques.

3.2 The dimension and basic facets of $\text{conv } S$

In this section we derive the dimension of the replacement polytope $\text{conv } S$ and investigate the inequalities used to define S in (4). Under natural assumptions we show that the replacement polytope is full-dimensional. Further, we show that all inequalities that are necessary in the original definition of the replacement polytope are facets of the same. Since the proofs are rather long, they are relegated to the Appendix.

LEMMA 11 *The polyhedron defined by*

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \dots, T - T_i, \quad i \in \mathcal{N}, \quad (7a)$$

$$-x_{it} \geq -1, \quad t = 1, \dots, T - 1, \quad i \in \mathcal{N}, \quad (7b)$$

is integral. \square

PROPOSITION 12 (dimension of the replacement polytope) *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then the dimension of $\text{conv } S$ is $(N + 1)(T - 1)$, that is, $\text{conv } S$ is full-dimensional.* \square

REMARK 13 The replacement polytope is *not* full-dimensional if $T_i = 1$ for some $i \in \mathcal{N}$, since it then holds that $x_{it} = z_t = 1$, $t = 1, \dots, T - 1$, for all $(x, z) \in \text{conv } S$. Letting $A^=$ denote the matrix corresponding to the equality subsystem of $\text{conv } S$, this would yield that $\text{rank} A^= \geq 2T - 2$ and, by Proposition 2, that $\dim(\text{conv } S) \leq (N - 1)(T - 1)$. However, the case that $T_i = 1$ is not interesting in practice since it would mean that component i must be replaced at every time step. \square

PROPOSITION 14 *If $T_i \geq 2$ for $i \in \mathcal{N}$, then each of the inequalities*

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell \in \{1, \dots, T - T_i\}, \quad i \in \mathcal{N},$$

defines a facet of $\text{conv } S$. \square

PROPOSITION 15 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities*

$$x_{it} \leq z_t, \quad t = 1, \dots, T-1, \quad i \in \mathcal{N},$$

defines a facet of $\text{conv } S$. □

PROPOSITION 16 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities*

$$z_t \leq 1, \quad t = 1, \dots, T-1,$$

defines a facet of $\text{conv } S$. □

PROPOSITION 17 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities*

$$x_{kt} \geq 0, \quad k \in \mathcal{N} : T_k \geq 3, \quad t = 1, \dots, T-1,$$

defines a facet of $\text{conv } S$. □

REMARK 18 The inequalities in Proposition 17 do *not* define facets for $k \in \mathcal{N}$ such that $T_k \leq 2$ due to the following. If $T_k = 2$ then, for each $s \in \{1, \dots, T-2\}$, $x_{ks} = 0$ implies that $x_{k,s+1} = z_{s+1} = 1$ (likewise, $x_{k,T-1} = 0$ implies that $x_{k,T-2} = z_{T-2} = 1$) which yields that $\text{rank } A^\# \geq 2$, where $A^\#$ denotes the matrix corresponding to the equality subsystem of $\text{conv } S$. Letting $\tilde{F}_{k_s} = \{(x, z) \in \text{conv } S \mid x_{k_s} = 0\}$, it follows that $\dim \tilde{F}_{k_s} \leq (N+1)(T-1) - 2$, which implies that \tilde{F}_{k_s} is *not* a facet of $\text{conv } S$. □

Now, the set S is defined by the constraints

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \dots, T-T_i, \quad i \in \mathcal{N}, \quad (8a)$$

$$0 \leq x_{it} \leq z_t \leq 1, \quad t = 1, \dots, T-1, \quad i \in \mathcal{N}, \quad (8b)$$

$$x_{it}, z_t \in \mathbb{Z}, \quad t = 1, \dots, T-1, \quad i \in \mathcal{N}, \quad (8c)$$

and it follows from Propositions 14–17 that all of the inequalities necessary in the description of the set S define facets of $\text{conv } S$. A natural question then arises: Is $\text{conv } S$ completely described by the continuous relaxation of (8)? Unfortunately, this is not the case, which is shown by the following example.

EXAMPLE 19 (continuous relaxation) Consider a system with $N = 2$, $T_1 = 3$, $T_2 = 4$, and $T = 5$. Then the problem to

$$\begin{aligned} \text{minimize} \quad & x_{11} + x_{12} + 2x_{13} + x_{14} + x_{21} + 100x_{22} + 100x_{23} + x_{24} \\ & + 10z_1 + 10z_2 + z_3 + 10z_4 \\ \text{subject to} \quad & (8), \end{aligned}$$

has the optimal solution

$$(x_{11}, x_{12}, x_{13}, x_{14}) = (0, 0, 1, 0), \quad (9a)$$

$$(x_{21}, x_{22}, x_{23}, x_{24}) = (1, 0, 0, 0), \quad (9b)$$

$$(z_1, z_2, z_3, z_4) = (1, 0, 1, 0), \quad (9c)$$

with objective function value 14. However, if we relax the integrality requirements, we get the optimal solution

$$(x_{11}, x_{12}, x_{13}, x_{14}) = (0.5, 0, 0.5, 0.5), \quad (10a)$$

$$(x_{21}, x_{22}, x_{23}, x_{24}) = (0.5, 0, 0, 0.5), \quad (10b)$$

$$(z_1, z_2, z_3, z_4) = (0.5, 0, 0.5, 0.5), \quad (10c)$$

with objective function value 13.5. Hence the convex hull of feasible solutions to (8) is not completely defined by the inequalities in (8). \square

3.3 A new class of facets: An example

Example 19 shows that the inequalities in (8) are not sufficient to describe $\text{conv } S$. However, according to the Propositions 14–17 all of the inequalities in (8) define facets of $\text{conv } S$. Since by Proposition 12 $\text{conv } S$ is full-dimensional (under reasonable assumptions) the minimal description of $\text{conv } S$ is unique. Therefore, all of the inequalities in (8) are necessary in the description of $\text{conv } S$.

To completely describe $\text{conv } S$ we need however also facets other than those in (8). We study the replacement polytope that arises in Example 19, that is, the convex hull of the set of all $x \in \{0, 1\}^{2 \times 4}$ and $z \in \{0, 1\}^4$ such that

$$x_{11} + x_{12} + x_{13} \geq 1, \quad (11a)$$

$$x_{12} + x_{13} + x_{14} \geq 1, \quad (11b)$$

$$x_{21} + x_{22} + x_{23} + x_{24} \geq 1, \quad (11c)$$

$$x_{11} \leq z_1, \quad (11d)$$

$$x_{12} \leq z_2, \quad (11e)$$

$$x_{13} \leq z_3, \quad (11f)$$

$$x_{14} \leq z_4, \quad (11g)$$

$$x_{21} \leq z_1, \quad (11h)$$

$$x_{22} \leq z_2, \quad (11i)$$

$$x_{23} \leq z_3, \quad (11j)$$

$$x_{24} \leq z_4. \quad (11k)$$

We denote this set by S_{ex} .

By using Chvátal–Gomory rounding (see [24, p. 210]) we construct a new valid inequality:

PROPOSITION 20 *The inequality*

$$z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4 \geq 2 \tag{12}$$

is a valid inequality for S_{ex} . □

We see that the inequality (12) is not satisfied by the optimal solution (10) to the continuous relaxation of the replacement problem in Example 19. In fact, if we add the inequality (12) to the continuous relaxation in Example 19 we get the optimal solution

$$\begin{aligned} (x_{11}, x_{12}, x_{13}, x_{14}) &= (0, 0, 1, 0), \\ (x_{21}, x_{22}, x_{23}, x_{24}) &= (1, 0, 0, 0), \\ (z_1, z_2, z_3, z_4) &= (1, 0, 1, 0). \end{aligned}$$

This is the solution in (9), that is, it is an optimal solution to the original problem! The valid inequality (12) in fact defines a facet of $\text{conv } S_{\text{ex}}$:

PROPOSITION 21 *The valid inequality (12) defines a facet of $\text{conv } S_{\text{ex}}$.* □

3.4 Conclusions

We have made an introductory study of the facial structure of the replacement polytope. It was shown that the replacement polytope is full-dimensional (if the life limits of the components are greater than or equal to two time steps) and found that the inequalities that are necessary to formulate the replacement problem also define facets of the replacement polytope. Unfortunately, these facets are not sufficient to represent the replacement polytope, as was shown by an example. By using Chvátal–Gomory rounding we have shown how to find a new class facets for the example problem. It is straightforward to generalize this class to any instance of the replacement problem. However, it still remains to investigate the strength of the continuous relaxation when the new class of facets is added to the replacement problem.

4 Towards a model for the stochastic optimization problem

Assume that $\{x(t), t \in [0, T]\}$ is an \mathbb{R}^p -valued stochastic process modeling the crack size of a component in an aircraft engine E during the life span $[0, T]$ of E . When we have several units E_1, \dots, E_n the \mathbb{R}^p valued processes $x_1(t), \dots, x_n(t)$ describe the crack sizes in the units.

During E_i 's life the unit will be serviced, at some times t_{i1}, t_{i2}, \dots . Of particular interest is therefore the time to service of a unit, which could be but is not necessarily determined by the time to failure, cf. the discussion in Sections 4.1 and 4.1.1. In general the times to service can be defined as

$$\begin{aligned} t_{i1} &= \inf\{t \geq 0 : x_i(t) \in C\}, \\ t_{i,k+1} &= \inf\{t > t_{ik} : x_i(t) \in C\} - t_{ik}, \text{ for } k \geq 1, \end{aligned} \quad (13)$$

where the critical region for the feature process C is a subset of \mathbb{R}^p . We make the assumption that C is independent of t . Thus we make the simplified assumption that the time dynamics of the feature process does not influence the reliability or the choice to make a repair. In a real physical situation this is typically too strong an assumption; C will depend on time but also on possibly other components in the feature process, e.g., the rate x'_i and acceleration x''_i of change in x_i : In this particular special case e.g., the first time to failure would rather be

$$t_{i1} = \inf\{t \geq 0 : (x_i(t), x'_i(t), x''_i(t)) \in C(t)\},$$

where $C(t)$ is a subset of \mathbb{R}^{3p} .

Furthermore C is assumed to be the same for all units i , i.e., we assume homogeneity between units. An interesting first question is whether C can be assumed to be the same for all cycles k in (13), corresponding to a renewal type condition. This can be tested, cf. Svensson [30]. The last assumption could however be loosened by allowing the critical regions to depend on the cycle, so C would be a function C_k of the cycle number k , cf. Section 4.1.2 in the sequel.

The main objective of Section 4.1 is the transfer of structural information: There are two different types of engines, I and II . These are structurally different, have different types of laws governing their features, their features are observed in a slightly different manner and most importantly the amount of data differs significantly; for type I there exists a large data set whereas for type II the amount of data is limited. However, the more interesting engine type is the type II engine, being a replacement of the type I engine. The approach to addressing this is by the use of transfer of structural information from type I to type II . The underlying assumption behind

this approach is that the same mathematical model describes the distributions of the lives for both engine types. Let $x(t)$ be a random element (a real number or a finite-dimensional vector) with unknown distribution $F = F(\cdot; \theta)$ parameterized by a parameter $\theta \in \Theta \subset \mathbb{R}^s$, with $s < \infty$. The distribution for type *I* engines is denoted F_I and that of type *II* engines by F_{II} . We assume that the distribution functions are completely unknown except for the parameter values θ . A natural assumption on the relation between F_I and F_{II} are that they belong to the same parametric class of distributions $\mathcal{P} = \{F(\cdot; \theta) : \theta \in \Theta\}$, and that only the values of the parameters θ_I and θ_{II} differ; this means that $F_I = F(\cdot; \theta_I)$ and $F_{II} = F(\cdot; \theta_{II})$.

In Section 4.1.4 we use a more specific modeling of the lives: The observations of the feature processes are of two main types. The first and more crude type of data consists of times to service, t_{ik} , for the units, with or without the corresponding feature values, $x_i(t_{ik})$. The second and finer type of data consists of repeated measurements in the same cycle k of the feature process $x_i(s_{i1}), x_i(s_{i2}), \dots$, with time points s_{ij} possibly passing the time to repair so that possibly $s_{ij} > t_{ik}$ for some j, k ; this last possibility is a necessity if we are to make inference on the time to repair for this type of data. Thus assume that the feature process $x(t)$, which in our application is crack size in a component, is a stochastic process with a distribution $F_x = F_x(\cdot; \theta_x)$, with unknown parameter θ_x . Typically, this feature process is not completely observed, but rather it is observed with some measurement error ϵ_i , with distribution $F_\epsilon(\cdot; \theta_\epsilon)$ at times s_{ik} , so that observations consist of $g(x(s_{ik}), \epsilon_i)$, for some function g , which therefore has the distribution $F(\cdot; \theta_x, \theta_\epsilon)$. Let C be some critical region for x such that the component fails the first time x hits C , so that $t_{ik} = \inf\{t : x(t) \in C\}$ is the failure time. Using the form of x it is possible to obtain the distribution $F_{t_{ik}} = F_{t_{ik}}(\cdot; \theta_x, \theta_\epsilon)$ of t_{ik} . We then treat the parameter θ as a nuisance parameter, and t_{ik} as the interesting random variable for which we want to obtain the distribution based on the data $g(x(s_{ik}), \epsilon_i)$; this is possible using likelihood techniques and in particular predictive profile likelihood techniques, cf. Mathiasen [22] and Bjornstad [8].

The resulting estimators of life distributions are continuous. However, since the optimization model only treats discrete data it is necessary to make a discretization of the distributions, using as few points of support as possible for the resulting discrete distribution, in order to keep the complexity of the corresponding optimization problem within tractable limits; this is the topic for Section 4.2.2. Finally, we draw some conclusions and discuss the implications of the stochastic modeling for the optimization problem.

4.1 Structural models for time between repairs

Assuming that the only data at hand are the times between repairs, it is of interest to study parametric models that describe these well, with a view of transformation of structural information from type *I* to type *II* engines. We will study the fit of two classes of models: non-stationary renewal processes (NSRP) and non-homogeneous Poisson processes (NHPP).

The data consists of times between replacements for three details in the flameholder component. For every detail k we observe a sequence $t^k = \{t_1^k, \dots, t_j^k\}$ of times between repairs where the last observation is possibly (right-) censored, meaning that for the last observation we may have the information that the time to next replacement is longer than the time observed. We model different details as independent, that is, the corresponding random vectors t^k and t^i are independent if $k \neq i$.

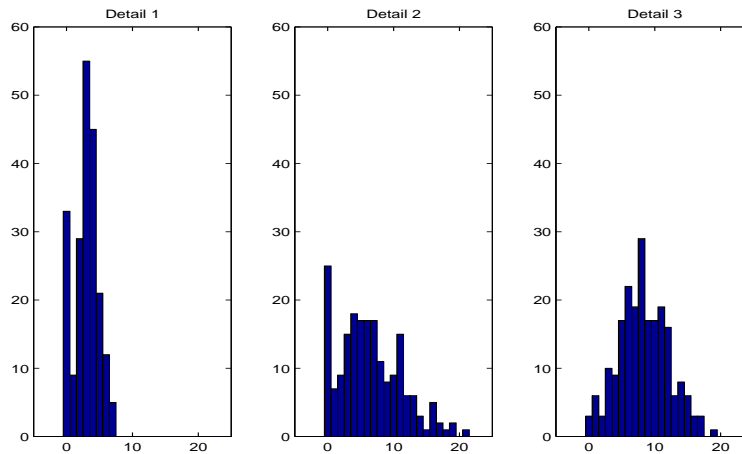


Figure 3: *Histogram on repaired components of each detail. The horizontal axes correspond to the number of repairs and the vertical ones to the number of component individuals that have been repaired the corresponding number of times.*

Assume that we have an arbitrary but fixed component. We introduce next the two types of stochastic processes we use to analyze the data: Let t_j be the time between the $j - 1$ 'th and the j 'th repair, and let $F_j(t) = P(t_j \leq t)$ be the corresponding distribution function. Let $N(t) = \#\{t_j \leq t\}$ be the corresponding counting process that counts the number of events that have occurred by time t .

DEFINITION 22 (renewal process) *An independently but not necessarily equally dis-*

tributed sequence $\{t_j\}_{j \geq 1}$ is called a non-stationary renewal process (NSRP); it is called stationary if $F_j \equiv F$ for all j and some F .

The inference problem consists of finding appropriate functions F_j and assessing whether in fact $F_j \equiv F$.

DEFINITION 23 (Poisson process) Let $\{N(t) : t \geq 0\}$ be a counting process with intensity function $w(t)$. If the process has independent and Poisson distributed increments it is called a non-homogenous Poisson process (NHPP); if $w(t) \equiv w$ is a constant function the process is a homogeneous Poisson process. \square

The NHPP is an appropriate model when one can suspect that there is dependence between repair times.

4.1.1 Survival analysis

To describe the distribution of the times between repairs $t_i = t_{i,k+1} - t_{i,k}$ for a unit E_i , we introduce the survival function $S_i(t) = P(t_k > t) = 1 - F_i(t)$ for unit i . The hazard function $h_i(t)$ is defined as $h(t) = -S'(t)/S(t)$, if S is absolutely continuous, and is interpretable as the time dependent conditional failure rate of unit i via the (formal) relation $h_i(t)dt = P(t_i \leq t + dt | t_i > t)$. The relations between the hazard and survival functions are given by

$$\begin{aligned} h(t) &= -\frac{d}{dt} \log S(t), \\ S(t) &= \exp\left(-\int_0^t h(u) du\right). \end{aligned} \tag{14}$$

For a particular unit that is active, meaning that its feature process x_i is not stopped, the last observation is possibly incomplete in that the unit does not yet satisfy the criteria for repair. In this case the unit's last observation time is censored, i.e., it is only known that the time until failure is larger than the observed time. Thus the data consists of pairs (t_{ik}, δ_{ik}) with t_{ik} the observed k 'th time for unit i and δ_{ik} an indicator variable for whether the observed time is a failure or a censoring. One typically assumes that censoring is due to other mechanisms than the ones governing the failure time, cf. Andersen et al. [2].

Assume that we have (possibly right censored) data (t_i, δ_i) . The standard estimator

of the survival function for right censored data is the Kaplan–Meier estimator

$$S_n(t) = \prod_{t_i \leq t} \left(1 - \frac{\delta_i}{Y_n(t_i-)}\right) = \prod_{u \leq t} \left(1 - \frac{dN_n(u)}{Y_n(u-)}\right), \quad (15)$$

where $Y_n(t) = \sum_{i=1}^n 1\{t_i > t\}$ are the number at risk at time t , and $N_n(t) = \sum_{i=1}^n 1\{t_i \leq t, \delta_i = 1\}$ are the number that have failed by time t . The hazard can not be estimated directly from data, since it is a (conditional) density, and neither can the density f . However, the integrated hazard function $H(t) = \int_0^t h(u) du$, can be estimated by the Nelson–Aalen estimator

$$H_n(t) = \sum_{t_i \leq t} \frac{\delta_i}{Y_n(t_i-)} = \int_0^t \frac{1}{Y_n(u-)} dN_n(u),$$

Using kernel smoothing techniques it is possible to estimate the hazard function as

$$h_n(t) = \int_0^\infty \frac{\frac{1}{h} k\left(\frac{t-u}{h}\right)}{Y_n(u-)} dN_n(u), \quad (16)$$

where k is a positive function integrating to one, typically with compact support (e.g., $[-1, 1]$) and h is the so called bandwidth.

The repeated measurements for one unit E_i are typically so few that it is not feasible to make separate inference for each individual unit based on data only from that unit. One approach could be to disregard possible differences between units and treat the gathered data as coming from one homogenous unit E . In this instance $S_i = S$ and $h_i = h$ are not assumed to depend on what particular unit E_i is being observed.

4.1.2 Model fit

Given a sample t_1, \dots, t_n with unknown distribution F we make a formal test of the hypothesis $H : F = F_0$. Since we want to use an adequate model to make a prediction of the time when the component fails, with the prediction of the expected failure time according to the model as the relevant prediction; the test statistic

$$M = \frac{1}{n} \sum_{i=1}^n (E_{F_0}(T_i) - t_i)^2,$$

seems reasonable.

For the NSRP model it is straightforward to calculate M since T_i is Weibull distributed and $E[T_i]$ then has a known parametric form. Table 1 shows the observed M and expected M_e squared error for the four first repairs of details 1, 2, and 3.

Table 1: Prediction error of the NSRP model for details 1, 2, and 3 for the four first failures and total time to failure.

Detail	$10^4 \times$	Failure 1		Failure 2		Failure 3		Failure 4		Average	
		M	M_e	M	M_e	M	M_e	M	M_e	M	M_e
1		3.28	3.41	1.92	2.12	1.57	1.89	1.61	2.50	2.12	2.45
2		2.02	2.02	0.06	0.06	0.06	0.06	0.06	0.06	0.34	0.34
935 3		7.20	7.45	0.22	0.21	0.29	0.26	0.28	0.26	1.05	1.08
329											

As is evident M_e is mostly bigger than M , which may be a consequence of the fact that the true distribution has shorter tails than the Weibull distribution.

In the NHPP model the prediction at time t_0 of the time to the next failure T is given by

$$E[T] = \int_0^\infty e^{-[W(t_0+t) - W(t_0)]} dt.$$

where W is the cumulative intensity

$$W(t) = \int_0^t w(u) du$$

It is not possible to estimate W for values of t larger than the largest observed life without assuming parametric models; one remedy for calculating $E(T)$ is to stop integration at the largest observation. An alternative method, given in Svensson [30], is exact if W is linear and an approximation if W is close to linear, cf. Table 2.

Table 2: Prediction error of the NSRP model for details 1, 2, and 3 in the four first failures and total time to failure.

Detail	$10^4 \times$	Failure 1	Failure 2	Failure 3	Failure 4	Average
		M	M	M	M	M
1		4.71	1.89	1.65	1.43	2.39
2		2.23	0.42	0.28	0.16	0.43
3		9.77	0.94	0.88	0.72	1.60

Table 3 shows that the NSRP model seems to be a better model for this data set.

There are two different repair stations: repair stations within close range of the aircraft A and the main central repair station B. To investigate if there is a difference

Table 3: Relative error for the two models.

Detail	$M_{\text{NSRP}} \times 10^4$	$M_{\text{NHPP}} \times 10^4$	$\frac{M_{\text{NHPP}}}{M_{\text{NSRP}}}$
1	2.12	2.39	1.13
2	0.34	0.43	1.26
3	1.05	1.60	1.48

between the repairs performed at A and B respectively we estimate the mean time to repair for the three details but distinguish between repair A and B. In Figure 4 this is shown for the first five repairs with 95% confidence interval for the mean. There seems to be a difference between the repair A and B, at least for details 1 and 3, cf. also Table 4.

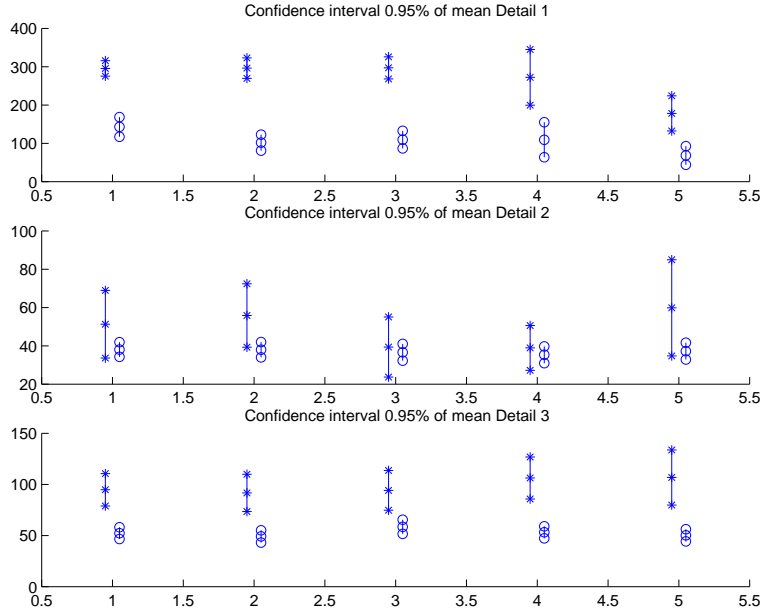


Figure 4: Confidence interval on expected time to failure after repair number 1 to 5. Stars represent repair type B and rings represent repair type A.

Next we test whether F_n depends on n . In Table 4 estimates of θ, α , of the mean of Weibull distribution μ_w , and of the mean of non-parametric distribution μ , are

Table 4: The parameters θ and α and the mean μ_w of the Weibull distribution, and the non-parametric estimated mean μ .

Detail	Failure no.	Repair Type	θ	α	μ_w	μ
1	1	-	554	3.09	496	498
1	2	A	145	1.48	131	131
1	3	A	159	1.46	144	142
1	4	A	109	1.26	101	102
1	5	A	125	1.98	110	110
1	6	A	119	1.37	109	109
1	2	B	359	3.00	321	323
1	3	B	330	3.14	295	295
1	4	B	335	2.83	298	296
1	5	B	331	3.69	299	297
1	6	B	297	3.31	266	272
2	1	-	355	2.36	315	315
2	2	A	45	1.76	40	40
2	3	A	43	1.69	38	38
2	4	A	43	1.76	38	38
2	5	A	41	1.69	37	37
2	6	A	40	1.76	36	35
2	2	B	41	2.20	36	36
2	3	B	52	1.04	51	51
2	4	B	62	1.64	55	56
2	5	B	37	0.84	41	39
2	6	B	44	2.29	39	39
3	1	-	589	1.99	522	526
3	2	A	76	1.77	67	67
3	3	A	58	1.61	52	52
3	4	A	54	1.43	49	49
3	5	A	65	1.49	58	58
3	6	A	59	1.66	53	53
3	2	B	101	1.91	90	90
3	3	B	105	1.47	95	95
3	4	B	98	1.27	91	92
3	5	B	103	1.39	94	94
3	6	B	119	1.68	106	106

shown for the time to first failure and for the following five times to failure after repair for details 1, 2, and 3. We notice that the estimates of θ and α do not change

very much between successive repairs if we disregard new components, which makes it natural to suggest a model with the same distribution for T_i , $i \geq 2$. Thus there seems to be two classes of repairs: repair of new components and repair of old ones.

An interesting hypothesis is whether the time to next repair decreases with the number of repairs. To test this we use the model

$$F_{T_n}(t) = 1 - e^{-\left(\frac{1}{\theta p^n} t\right)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0, p > 0), \quad (17)$$

where n is the repair number, implying that the expected time to failure after repair number n is

$$E[T_n] = \theta p^n \cdot \Gamma\left(\frac{1}{\alpha} + 1\right),$$

and $p < 1$ thus indicates aging. Maximum likelihood estimates of the parameters (θ, α, p) are shown in Table 5 together with 95% confidence interval over the true parameter p based on profile likelihood. It seems that if we use $p = 1$ the resulting error is very small; no aging parameter is therefore necessary.

Table 5: *Parameters in modified Weibull distribution when T_i , $i \geq 2$ have the same distribution and a 95% confidence intervals over the parameter p .*

Detail	Repair type	θ	α	p	95% confidence interval
1	A	164	1.43	0.92	(0.8412 , 1.0146)
1	B	371	3.04	0.95	(0.9205 , 1.0026)
2	A	45.7	1.79	0.98	(0.9666 , 0.9902)
2	B	47.0	1.36	1.01	(0.9205 , 1.0026)
3	A	61.8	1.51	0.99	(0.9794 , 1.0058)
3	B	101.9	1.55	1.01	(0.9846 , 1.0426)

4.1.3 Parametric modeling and likelihood based inference

A refinement of the above inference approach is possible if there were variables that would influence the distribution of the time between failures. Thus assume that for each unit E_i we are able to measure a vector of covariates $z_i = (z_{i1}, \dots, z_{ip})$ and that these affect the survival and hazard functions so that $S_i(t) = S(t; z_i)$ and $h(t) = h(t; z_i)$. The standard approach to modeling for inclusion of covariates is the Cox proportional hazards model. In this one makes the assumption

$$h(t; z_i) = h_0(t) e^{\beta_1 z_{i1} + \dots + \beta_p z_{ip}},$$

with h_0 an unspecified function, the so called baseline hazard, that is the hazard for an individual that has no covariate load, and $\beta = (\beta_1, \dots, \beta_p)$ a vector of unknown

parameters that determine the size of the effect of the covariate z_{ip} on the hazard function h and thus on the survival function S . However, in our case the potentially interesting covariates are highly correlated with the lives, and therefore not amenable to analysis.

Another approach to modeling the underlying causes for the distribution of lives is via a physical model; this is more informative than a Cox regression model, since the latter does not have any physical justification; the great success of the latter depends to a large extent on the generalizability of and ease of calculation in the model.

Assuming that the feature process $\{x(t) : t \geq 0\}$ follows a particular form, e.g., based on physical modeling, it is possible to derive formal expressions for the distribution of the time between repairs.

4.1.4 The one dimensional case: Crack length modeling

In particular let $\{x(t) : t \geq 0\}$ be a real valued stochastic process that describes the growth of a crack in a component that is crucial for the reliability of a unit. Assume that a failure occurs when the crack grows past a critical point c_p , so that the critical region is $C = [c_p, \infty)$.

The stochastic model we use for crack growth is

$$a(t) = \begin{cases} a_0, & \text{if } t < S, \\ a_0 + C(t - S)^b, & \text{if } t \geq S, \end{cases} \quad (18)$$

where $a(t)$ is the crack length at time t , C and S are two stochastic variables and $b > 0$ is a parameter, cf Svensson [31]. The parameter b is assumed constant for all cracks, while C and S are particular to the crack in question: Thus for crack number i we have outcomes c_i, s_i of C, S and the crack is assumed to follow the model above with C, S replaced by c_i, s_i .

However, we do not observe the cracks directly but rather with some measurement error: Thus for each crack $i = 1, \dots, k$ we have n_i observations $x_{ij} = x_i(t_{ij})$ at times $t_{i1} < \dots < t_{in_i}$. The observations are assumed to follow the model

$$X_i(t_{ij}) = \begin{cases} a_0 & \text{if } t_{ij} < s_i, \\ \max \{a_0, a(t_{ij}) + \varepsilon_{ij}\} & \text{if } t_{ij} \geq s_i, \end{cases}$$

with ε_{ij} being a sequence of independent $N(0, \sigma_\varepsilon^2)$ distributed random variables. Thus the distribution of $X_i(t_{ij})$, is conditionally on (c_i, s_i) , a mixture of a discrete

and continuous distribution, the discrete random variable having a point mass a_0 at s_i .

Let $\theta = (\psi, \sigma_\varepsilon^2, b)$. For crack i we want to find the the distribution of the time

$$T = \inf\{t : a(t) \geq a_{\max}\}.$$

Assume that we are given previous observations \mathbf{x}_i of crack i , and assume the parameters θ of the model are known to us.

Using model (18), it is possible to get an expression for the conditional distribution and density

$$\begin{aligned} P(T \leq t \mid \mathbf{X}_i = \mathbf{x}_i), \\ f_{T|\mathbf{X}_i}(t|\mathbf{x}_i; \theta), \end{aligned} \quad (19)$$

which via Bayes' formula implies

$$f_{(C,S)|\mathbf{X}_i}(c, s \mid \mathbf{x}_i; \theta) = \frac{f_{\mathbf{X}_i|C,S}(\mathbf{x}_i \mid c, s; b, \sigma_\varepsilon^2) f_{C,S}(c, s; \psi)}{f_{\mathbf{X}_i}(\mathbf{x}_i; \theta)}. \quad (20)$$

Using the model descriptions it is possible to obtain expressions for all factors in (20), and thus for the parameters in the crack model given observations of cracks.

Let A be the crack length of crack i at a fixed time t . Similarly we have

$$f_{A|\mathbf{X}_i}(a, \mid \mathbf{x}_i; \theta) = \frac{f_{\mathbf{X}_i|A}(\mathbf{x}_i \mid a; \theta) f_A(a; \theta)}{f_{\mathbf{X}_i}(\mathbf{x}_i; \theta)}. \quad (21)$$

and it is possible to obtain expressions for all factors in 21.

The likelihood for the parameters given data of cracks $\underline{\mathbf{x}} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is

$$L(\theta) = f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta) = \prod_{i=1}^n f_{\mathbf{X}_i}(\mathbf{x}_i; \theta), \quad (22)$$

In order to take the parameter uncertainty in account we use a profile likelihood approach, c.f Pawitan [25], and define the predictive profile likelihood for T given x_i

$$\tilde{L}(t \mid \mathbf{x}_i; \underline{\mathbf{x}}_{-i}) = \sup_{\theta} f_{T|\mathbf{X}_i}(t \mid \mathbf{x}_i; \theta) f_{\underline{\mathbf{x}}_{-i}}(\underline{\mathbf{x}}_{-i}; \theta). \quad (23)$$

It is common to normalize (23) to obtain a density function

$$\tilde{f}_{T|\mathbf{X}_i}(t \mid \mathbf{x}_i) = \frac{\tilde{L}(t \mid \mathbf{x}_i; \underline{\mathbf{x}}_{-i})}{\int \tilde{L}(t \mid \mathbf{x}_i; \underline{\mathbf{x}}_{-i}) dt},$$

and use it for inference of T given \mathbf{x}_i .

If the number of cracks k is large the predictive profile likelihood in (23) is numerically demanding to compute. We then suggest to ignore the uncertainty in the parameter estimation and use $\theta = \hat{\theta}$ for all times t . The likelihood then becomes

$$L(\hat{\theta}, t | \mathbf{x}_i; \underline{\mathbf{x}}_{-i}) = f_{T|\mathbf{X}_i}(t | \mathbf{x}_i; \hat{\theta}) f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i}; \hat{\theta}) = f_{T|\mathbf{X}_i}(t | \mathbf{x}_i; \hat{\theta}) K,$$

where K is a constant eliminated by a normalization. The predictive profile likelihood for inference on the crack length at a specific time is given by

$$\tilde{L}(A | \mathbf{x}_i; \underline{\mathbf{x}}_{-i}) = \sup_{\theta} f_{A|\mathbf{X}_i}(a | \mathbf{x}_i; \theta) f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i}; \theta). \quad (24)$$

4.1.5 Case study

We use the above model to make predictions on crack growth in a low pressure turbine nozzle component, for a small data set. We present estimates of the joint distribution of C and S . and an illustration of the difference of estimating the remaining life with and without taking the uncertainty in the parameter estimation into account.

The data available is from pri-engines. Pri-engines are engines that are used extensively so they accumulate a large number of flight hours and flight missions. The engines have been observed every 200 flight hours.

We first determine the joint distribution of C and S . Figure 5 illustrates the cracks and the crack model (18) fitted to the cracks with a least square method. From each picture we get an observation of S (censored if no crack was detected) and an observation of C if a crack was detected.

Using a similar procedure for all cracks (more than the four in Figure 5) indicates that S and C are uncorrelated, which makes it feasible to assume independence between S and C . Furthermore, by examining the empirical distribution of S and C we find that the log normal distribution gives a reasonable fit; we therefore assume that both S and C are log normally distributed with parameters μ_s, σ_s and μ_c, σ_c , respectively.

4.1.6 Model illustration

Assume that we have observed the cracks on the three first components and we want to predict when the crack on the fourth component reaches the length $a_{\max} = 30$ mm, that is to find the distribution of T given the observations of component four.

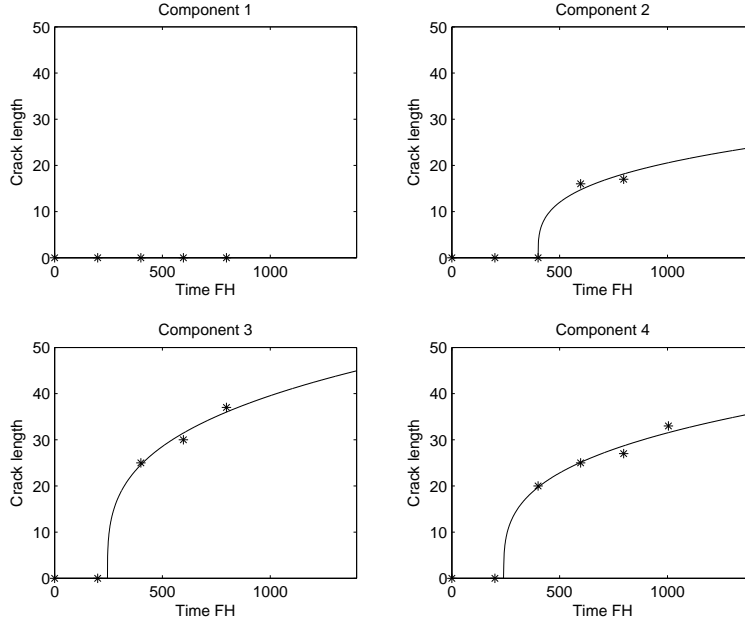


Figure 5: Model (18) with $a_0 = 0$ fitted to the four cracks.

It is also possible to update the distribution of T each time we get new observations, at 200, 400, 600 and 800 flight hours.

Assume that we know from experience that $\sigma_\varepsilon = 1$ mm, and the other parameters are unknown. We use the observations from components one, two and three to estimate the remaining parameters $(\mu_s, \sigma_s, \mu_c, \sigma_c, b)$ using equation (22). First we calculate the distribution of T using (19), ignoring the uncertainty in the parameter estimation. The solid line in Figure 6 illustrates the distribution of T when ignoring parameter estimation uncertainty and $\theta = \hat{\theta}$.

Using the profile likelihood approach (23) we see how much the uncertainty in the parameter estimates affects the results. In Figure 6 the line with stars is a plot of the distribution of T when we consider the uncertainty of the parameters, the stars indicating where the distribution has been calculated.

The crack length distribution at a fixed time is obtained similarly. Figure 7 shows the crack growth over time given the information in our observations.

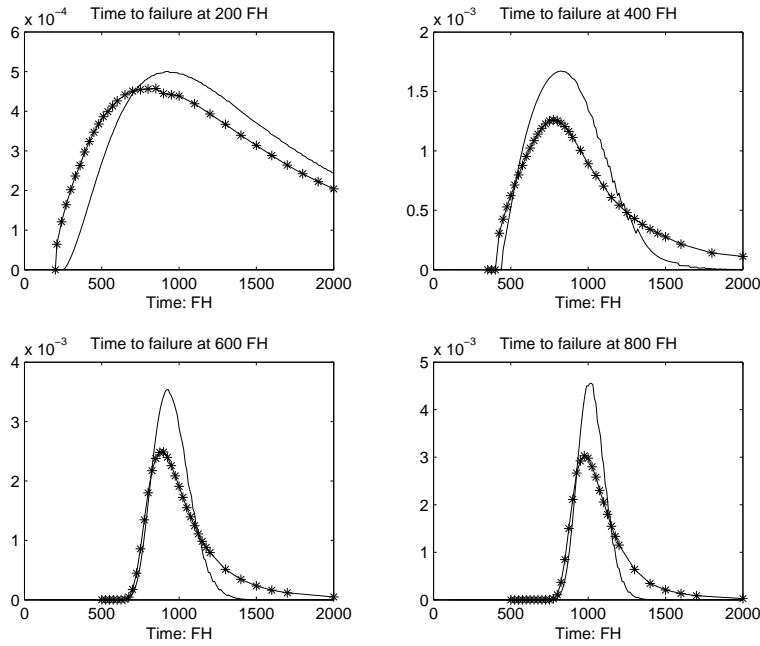


Figure 6: *Distribution of the time when the crack of component four will reach 30mm both considering uncertainty in parameter estimation (line with stars) and without uncertainty (solid line). The distribution is updated with the observations at 200 flight hours (upper left), 400 flight hours (upper right), 600 flight hours (lower left) and 800 flight hours (lower right).*

Integrating the function in Figure 7 with respect to crack length we get marginal function of value 1 for all times. The high values in the upper part of the pictures is the point mass that indicates that the crack length is below size a_0 . In the case of one observation, the upper left picture, we can observe how the probability of a crack length of length a_0 decreases as time increases. At 0 FH this probability is zero. The observation that we have no crack at 200 FH gives a very slim chance that there would be a crack at this time, hence the probability is almost one. In the other pictures in Figure 7 we get a feeling that even if we are fairly certain of the crack length at a fixed time the distribution of when the crack reaches a specific crack length will have a large variance.

If we combine the probability model with the profile likelihood based approach we get a comparison of the effect of the uncertainty in the parameters on the distribution

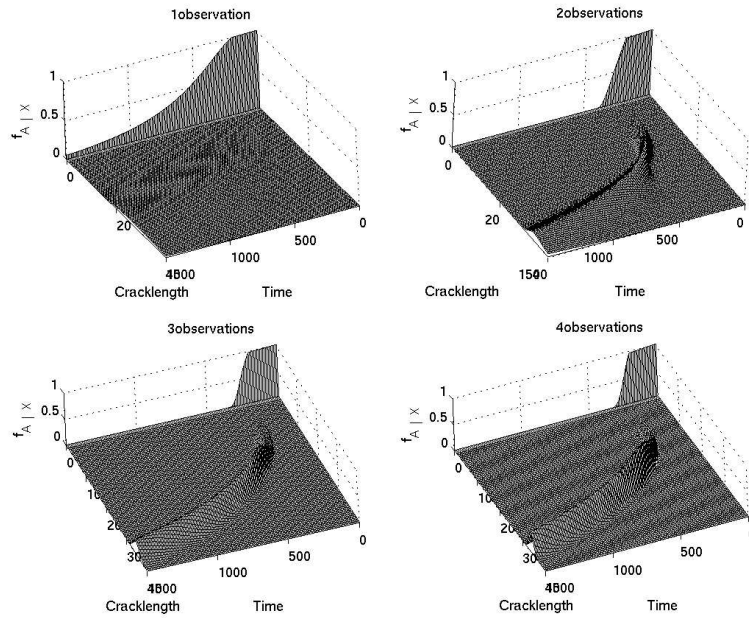


Figure 7: *Three dimensional illustration of how the distribution of the crack length at a fixed time of component four changes over time. The distribution is updated each time a new observation is made: first observation at 200 flight hours (upper left), second observation at 400 flight hours (upper right), third observation at 600 flight hours (lower left) and fourth observation at 800 flight hours (lower right). The values at crack length 0 corresponds to the probability point mass that there are no cracks.*

of crack length at fixed time. This is plotted in Figure 8 for times 500 FH and 1000 FH.

4.2 An optimal discretization of a continuous distribution

We will treat the case with one stochastic component, which can be extended to several stochastic components. The life U of a new stochastic component is modeled with a distribution \check{G} and the remaining life of a functioning stochastic component

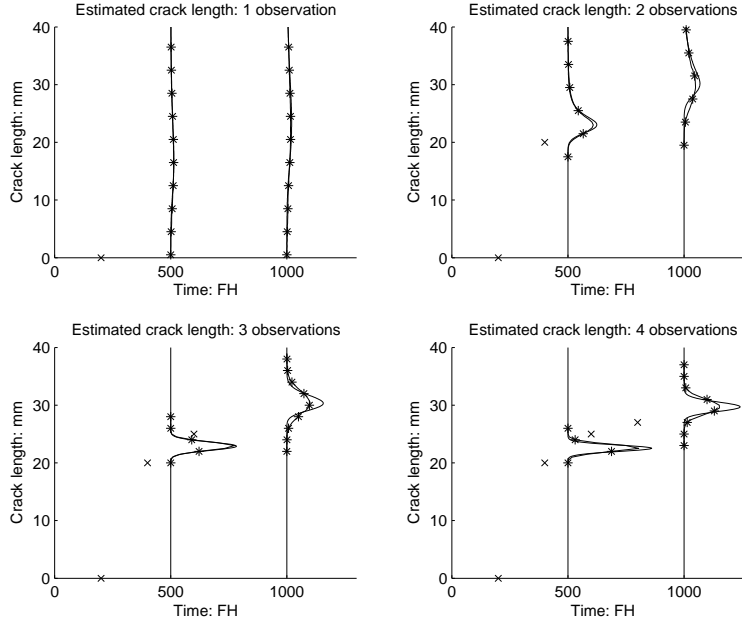


Figure 8: Distribution of the crack length at times 500 and 1000 flight hours of component four both considering uncertainty in parameter estimation, line with stars, and without uncertainty, solid line. The distributions are updated with the observations at 200 flight hours (upper left), 400 flight hours (upper right), 600 flight hours (lower left) and 800 flight hours (lower right). The observations are marked in the pictures with "x".

of age of u_0 is modeled with a distribution G defined by

$$G(u) = \frac{\check{G}(u + u_0) - \check{G}(u_0)}{1 - \check{G}(u_0)}.$$

We assume that $\check{G}'(u) > 0$ if $u > 0$ and that U is a non-negative random variable.

There are two main simplifications made: The first is that in the sequence of life distributions for the stochastic component only the first life distribution is modeled as a random variable, the remaining life distributions are replaced by a functional of the distribution such as the expected value or the median. The second simplification necessary is due to the fact that the optimization model does (at least in practice) not work with continuous time, since it is defined using discrete time points.

In the sequel we distinguish between the first and second stage models. The first stage takes into consideration all possible future events, while the second stage model contains one model for each future event.

Overall we wish to minimize the expected cost of maintaining the engine during a fixed time period containing T equidistant time points. Thus components are only allowed to be replaced at these time points. The life limits of the deterministic components are T_i , $i \in \mathcal{N}$, as defined in Section 2. The time to the first failure of the stochastic component is modeled with the distribution G : The life U of the stochastic component currently in the engine is defined as $\tilde{\tau}(u) \in \{1, \dots, T-1\}$. The life of each replacing stochastic component is defined as $\tau \approx E_G U$ (see Altenstedt [1]) such that $\tau \in \{1, \dots, T-2\}$ (if $\tau \geq T-1$ the stochastic component is replaced at most once during the time horizon T). In addition to the costs defined in Section 2, c denotes the cost for a replacing stochastic component.

The binary variables x_{it} , representing the replacement of the deterministic components, and z_t , representing maintenance occasions, are defined as in Section 2.1. The binary variables s_t are defined as

$$s_t = \begin{cases} 1, & \text{if the stochastic component is replaced at time } t, \\ 0, & \text{otherwise,} \end{cases} \quad t \in \{1, \dots, T-1\}.$$

The first stage binary variables are x_{i1} for the deterministic components $i \in \mathcal{N}$, s_1 for the stochastic component, and z_1 for the maintenance occasion. We introduce the replacement strategy vector $\mathbf{x}_1 = (x_{11}, \dots, x_{N1}, s_1, z_1)$ and define an optimal replacement strategy for the first stage as

$$\hat{\mathbf{x}}_1 \in \underset{\mathbf{x}_1 \in \mathbb{B}^{N+2}}{\operatorname{argmin}} F(\mathbf{x}_1),$$

where

$$F(\mathbf{x}_1) = \int_0^\infty f(\mathbf{x}_1, u) dG(u) = E_G[f(\mathbf{x}_1, U)]. \quad (25)$$

The second stage function $f(\mathbf{x}_1, u)$ represents the cost for the maintenance schedule conditioned that the replacement strategy for the first stage is fixed to \mathbf{x}_1 and that

the life of the stochastic component currently in the engine is u . It is defined as

$$\begin{aligned}
f(\bar{\mathbf{x}}_1, \tilde{\tau}(u)) = \min & \quad \sum_{t=1}^{T-1} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + cs_t + dz_t \right), & (26) \\
\text{subject to} & \quad \sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \dots, T - T_i, \quad i \in \mathcal{N}, \\
& \quad \sum_{t=1}^{\tilde{\tau}(u)} s_t \geq 1, \\
& \quad \sum_{t=\ell}^{\tau+\ell-1} s_t \geq 1, \quad \ell = \max\{2, \tilde{\tau}(u) - \tau + 1\}, \dots, T - \tau, \\
& \quad \sum_{t=\ell}^{\tau+\ell-1} s_t \geq s_{\ell-1}, \quad \ell = 2, \dots, \tilde{\tau}(u) - \tau, \\
& \quad \quad \quad \text{if } \tilde{\tau}(u) - \tau \geq 2, \\
& \quad x_{it} \leq z_t, \quad t = 1, \dots, T - 1, \quad i \in \mathcal{N}, \\
& \quad s_t \leq z_t, \quad t = 1, \dots, T - 1, \\
& \quad \mathbf{x}_1 = \bar{\mathbf{x}}_1, \\
& \quad x_{it}, s_t, z_t \in \mathbb{B}, \quad t = 1, \dots, T - 1, \quad i \in \mathcal{N}.
\end{aligned}$$

The computation of the second stage function (26) requires a discretization of the distribution G . Let $n \in \{1, \dots, T - 1\}$ and define $\kappa_n = \{k_1, \dots, k_n\} \subseteq \{1, \dots, T - 1\}$ such that $k_{j+1} > k_j$, $j \in \{1, \dots, n - 1\}$. An n -discretization G_n of the distribution G has the probability mass function

$$g_n(u) = \begin{cases} p_j, & \text{if } u = k_j, \quad j = 1, \dots, n, \\ 0, & \text{if } u \notin \kappa_n. \end{cases} \quad (27)$$

A number of methods for defining κ_n and p_j , $j = 1, \dots, n$, are described in Section 4.2.2.

4.2.1 Error measure

To reduce the solution time for the optimization requires a discretization with as few points of support as possible. The n -discretization (27) yields the replacement strategy

$$\hat{\mathbf{x}}_1^n \in \operatorname{argmin}_{\mathbf{x}_1 \in \mathbb{B}^{N+2}} F_n(\mathbf{x}_1), \quad (28)$$

with F_n defined as F in (25) with G replaced by G_n , so that

$$F_n(\mathbf{x}_1) = \sum_{j=1}^n f(\mathbf{x}_1, k_j) \cdot p_j.$$

As G_{T-1} is a distribution using the maximum number of support points in the model we compare the quality of G_n to that of G_{T-1} by introducing the error measure for the expected cost between the two discretizations as

$$e(G_n, G_{T-1}) = F_{T-1}(\hat{\mathbf{x}}_1^n) - F_{T-1}(\hat{\mathbf{x}}_1^{T-1}). \quad (29)$$

In order to get a small error we use the following result (cf. Svensson [32] for a proof).

THEOREM 24 *The error measure is bounded by the following:*

$$e(G_n, G_{T-1}) \leq 2 \cdot \sup_{\mathbf{x}_1 \in \mathbb{B}^{N+2}} |F_n(\mathbf{x}_1) - F_{T-1}(\mathbf{x}_1)| \leq C \cdot \sup_{u \in \mathbb{R}} |G_n(u) - G_{T-1}(u)|, \quad (30)$$

where $C \in \mathbb{R}_+$ is bounded. □

4.2.2 Discretization approaches

When approximating a discretization of $G(u), u \in [0, \infty)$ with $n \leq T - 1$ points of support, the following questions arise: 1) How many points of support to use? 2) Which points of support $\kappa_n \subseteq \kappa_{T-1}$ should we choose? 3) How should we place the probability mass? Answering the questions 2) and 3) simultaneously may lead to optimization problems that are as difficult to solve as the original problem. In Section 4.2.3 we try to answer question 1) using simulation. In Svensson [32] four different approaches were described: (i) Minimizing the supnorm distance, (ii) using means in brackets, (iii) minimizing the Wasserstein distance, and (iv) moment preserving discretization.

Method minimizing the supnorm-distance. A simple calculation (Svensson [32]) shows that

$$\sup_u |G_n(u) - G_{T-1}(u)| - \varepsilon \leq \sup_u |G_n(u) - G(u)| \leq \sup_u |G_n(u) - G_{T-1}(u)| + \varepsilon,$$

where $\varepsilon = \sup_u |G_{T-1}(u) - G(u)|$ is small when T is large.

We next describe two approaches to bound the supnorm-distance:

(a) If both the probabilities $\mathbf{p} = \{p_1 \dots p_n\}$ and the n support points $\kappa_n \subseteq \kappa_{T-1}$ are free, we define

$$(\hat{\mathbf{p}}, \hat{\kappa}_n) \in \operatorname{argmin}_{\mathbf{p}, \kappa_n \subseteq \kappa_{T-1}} \left\{ \sup_u |G(u) - G_n(u)| \right\}. \quad (31)$$

The solution is not unique since it only considers the greatest difference between G and G_n . (b) Instead of the above optimization we suggest the following approach:

1. Choose the first point of support as

$$k_1 \in \operatorname{argmin}_{u \in \kappa_{T-1}} \left\{ \left| G(u) - \frac{1}{2n} \right| \right\}.$$

2. Choose the i 'th point of support as

$$k_i \in \operatorname{argmin}_{u \in \kappa_{T-1} \setminus \{k_1, \dots, k_{i-1}\}} \left\{ \left| G(u) - \frac{2(1 - G(k_{i-1}))}{2(n - i + 1) + 1} - G(k_{i-1}) \right| \right\}. \quad (32)$$

Then choose probabilities as

$$\begin{aligned} p_1 &= \frac{G(k_1) + G(k_2)}{2}, \\ p_i &= \frac{G(k_i) + G(k_{i+1})}{2} - \frac{G(k_{i-1}) + G(k_i)}{2} = \frac{G(k_{i+1}) - G(k_{i-1})}{2}, \\ p_n &= 1 - \frac{G(k_n) + G(k_{n-1})}{2}, \end{aligned}$$

where $i = 1, \dots, n - 1$.

Bracket means method. Partition G into n intervals $\{[t_0, t_1], \dots, [t_{n-1}, t_n]\}$, where $t_j > t_{j-1}$, $j = 1, \dots, n$, $t_0 = 0$, and $t_n = \infty$, and assign the probability masses

$$p_j = \int_{t_{j-1}}^{t_j} dG(t), \quad j = 1, \dots, n, \quad (33)$$

to the time points

$$u_j = \frac{1}{p_j} \cdot \int_{t_{j-1}}^{t_j} t dG(t), \quad j = 1, \dots, n. \quad (34)$$

The intervals can be chosen in many ways, e.g., so that $p_j = 1/n$ for all j . In the case where the points of support have to be in the set κ_{T-1} , it is not always possible to find subintervals such that the mean in each subinterval corresponds to a time point in κ_{T-1} . We present two approximations.

Approach 1 Let $\mu_1(a, b)$ be a metric measuring the distance between the point sets a and b . Choose a subset $\kappa_n \subseteq \kappa_{T-1}$. Let $\tilde{\kappa}_n = \{\tilde{k}_1, \dots, \tilde{k}_n\} \not\subseteq \kappa_{T-1}$. The solution $\hat{\mathbf{p}} = \{p_1, \dots, p_n\}$ is obtained by finding times $\{t_i, \dots, t_{n-1}\}$ such that

$$\hat{\mathbf{p}} \in \underset{\mathbf{p} \in \phi}{\operatorname{argmin}} \mu_1(\kappa_n, \tilde{\kappa}_n) \quad (35)$$

where

$$\phi = \left\{ \mathbf{p} \in \mathbb{R}^n \left| p_j = \int_{t_{j-1}}^{t_j} dG(t), \tilde{k}_j = \frac{1}{p_j} \int_{t_{j-1}}^{t_j} t dG(t), \right. \right. \\ \left. \left. t_j > t_{j-1}, j = 1, \dots, n, t_0 = 0, t_n = \infty \right\}. \quad (36)$$

Solving (35) yields the discretization (27).

Approach 2 Let μ_{21} measure the distance between probability vectors and let μ_{22} measure the distance between point sets. Let $\mathbf{q} = \{q_1, \dots, q_n\}$ be a probability vector with desirable probabilities and let $\tilde{\kappa}_n = \{\tilde{k}_1, \dots, \tilde{k}_n\} \not\subseteq \kappa_{T-1}$. Then a discretization $\mathbf{p} = \{p_1, \dots, p_n\}$ with points of support $\kappa_n = \{k_1, \dots, k_n\}$ is obtained as the solution $(\hat{\mathbf{p}}, \hat{\kappa}_n)$ to the optimization problem

$$(\hat{\mathbf{p}}, \hat{\kappa}_n) \in \underset{\mathbf{p} \in \phi, \kappa_n \subseteq \kappa_{T-1}}{\operatorname{argmin}} [\mu_{21}(\mathbf{p}, \mathbf{q}) + \mu_{22}(\kappa_n, \tilde{\kappa}_n)], \quad (37)$$

where ϕ is defined in (36). The solution of (37) yields the discretization of the distribution as in (27).

Method minimizing the Wasserstein distance In Pflug [26] the following discretization was suggested:

$$p_j = \int_{a_{j-1}}^{a_j} dG(t), \quad j = 1, \dots, n,$$

where where $a_0 = 0$, $a_j = \frac{k_j + k_{j+1}}{2}$, $j = 1, \dots, n-1$, $a_n = \infty$, and $\{k_1, \dots, k_n\} = \kappa_n$ are the points of support of the given G_n , derived as the discrete distribution minimizing the Wasserstein distance.

Moment preserving method If there are no constraints on the points of support and the first $2n - 1$ moments of the distribution G are finite, then it is possible to create a discrete approximation with n points of support that correctly matches $2n - 1$ moments. Let

$$M_k = \int_{-\infty}^{\infty} u^k dG(u),$$

be the k 'th moment. The discretization can be obtained by searching u_j and p_j that satisfy

$$\sum_{j=1}^n p_j u_j^k = M_k, \quad k = 0, \dots, 2n - 1. \quad (38)$$

We must choose points from the set κ_{T-1} . In order to compare the methods we will use the same points as for the method that minimizes the sup-distance. Rewriting the problem it can be seen that one may need to remove moment constraints. Then the solution to the problem is not necessarily unique. In order to choose one solution we use a function z that represents some desired properties of the discretization (the function z can, e.g., be formulated to promote p_j of the same sizes) and solve

$$\hat{\mathbf{p}} \in \operatorname{argmax}_{\mathbf{p} \in \varphi} z(\mathbf{p}),$$

where

$$\varphi = \left\{ \mathbf{p} \in \mathbb{R}^n \left| \sum_{j=1}^n p_j u_j^k = M_k, \quad k = 0, \dots, n - 1 - m, \quad p_j \geq 0, \quad j = 1, \dots, n \right. \right\}$$

and m is the number of removed moment constraints.

4.2.3 Test results

We describe the life of the stochastic component with the parametrization,

$$\check{G}(u) = 1 - e^{-(\frac{1}{\theta}u)^\alpha}, \quad u \geq 0, \quad (39)$$

where $\theta > 0$ is the characteristic life and $\alpha > 0$ is the shape parameter. The maximum number of time steps is $T = 30$, the distance between the time steps equals one, and $\theta = 9$. Tests were made with $\alpha \in \{1, 2\}$.

We model the engine with two components, one stochastic and one deterministic. In each time step there are hence four alternatives: (1) Replace the deterministic component, (2) replace the stochastic component, (3) replace both components, or (4) do not replace any components. Optimal replacement alternatives were calculated using the optimization model (28) with $n \in \{1, \dots, 10\}$ points of support.

The points of support were chosen as in (32). The probabilities were chosen using the method that minimizes the sup-distance, the method that minimizes the Wasserstein distances, the method that preserves the moments, and the bracket meansmethod approach 1.

The best discretization possible had $T - 1 = 30$ points of support, one in every time point, for which an optimal replacement strategy was calculated according to (28) with $n = T - 1$. The difference between the two discretizations, using the error measure (29), was calculated. The parameters and remaining lives of the components used are shown in Table 6.

Table 6: *Parameters for the optimization model (28) and the distribution (39).*

Parameter	Alternatives
$\tilde{\tau}(u)$	4, 6
τ	6, 10
c	60, 70, 100, 130, 150
d	70, 100, 150
α	1, 2

These results indicate that the method that preserves moments seems worse and the methods that minimize the Wasserstein and the supnorm-distance seem better. Furthermore, it seems that using two points of support is worse than just using the expected value of the distribution. Also it seems that there is a large gain in going to three points of support and that there is not a large gain in going further to more than three points of support, if we disregard the moment method with $\alpha = 1$.

Further tests showed that the error measure decreases as α increases. With constant θ , the variance of the Weibull distribution (39) decreases with an increasing value of the parameter α . If the variance is large then it is difficult to get a good approximation of the distribution with just a few points of support.

Finally, in Svensson [32] it was shown, using a so called narrow scenario tree scheme, that many details were replaced even if they did not need replacement, which seems to be due to the approximation that only the first life of the stochastic component is treated as stochastic, and approximated with a discrete distribution, while the remaining lives are treated as deterministic with lives equal to the expected value under the true distribution.

4.3 Output and relation to the optimization problem

The results of the mathematical modeling of the time between failures consist of three main components:

The first is an empirical evaluation of the distribution of the observed times, with

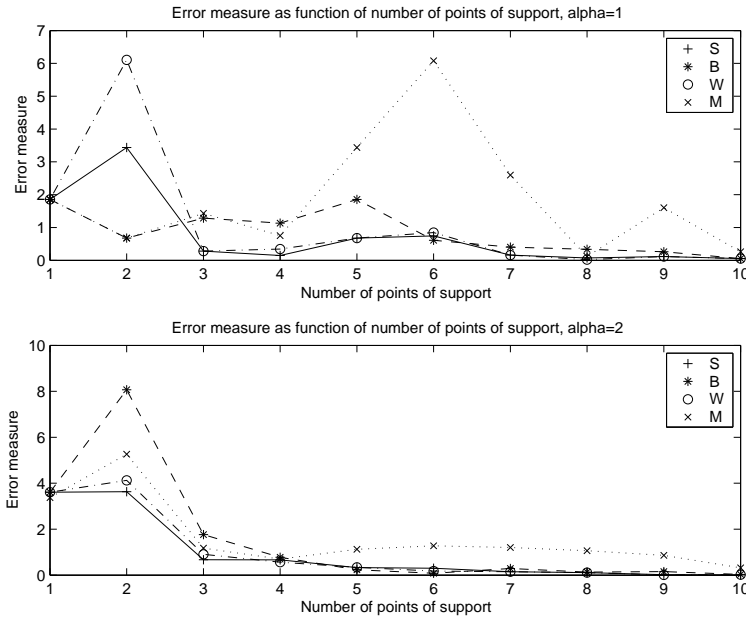


Figure 9: The vertical axis represents the mean error measure with parameters as in Table 6. The horizontal axis represents the number of support points. *S* means the method minimizing the sup-distance, *B* means the bracket method, *W* means the method minimizing the Wasserstein distance, and *M* means the moment preserving method.

respect to the problem of finding possible models that accurately describe the process of failure times. The main objective in this is to find a more narrow description of occurring distributions for the lives, that incorporates finite-dimensional parameters. The goal is to be able to transform structural information from one type of engine—for which there is a large set of empirical data—to another type of engine—for which the amount of data is more limited. Therefore, tests were made for two different types of stochastic processes, the non-homogenous Poisson process and the non-stationary renewal process, to see which was the better fitting.

The second part consists of physical modeling of, e.g., the growth of cracks with which one obtains more specific descriptions of the time until failure, which is defined as the first time the crack reaches a critical length. This part can be seen as a refinement of the standard approach when using more information.

Finally, since the optimization model for finding a replacement strategy is not devel-

oped to deal with probability distributions as inputs, but rather with discrete data, it was necessary to discretize the continuous distributions to discrete ones with only a few support points. Methods for the discretization were presented, using different strategies, and they were evaluated in a simple simulation study. Although the simulations were performed for a simple model they clearly indicate that increasing the number of support points can make a dramatic change in the cost savings. In the case studied, typically three support points seem adequate. This seems promising for the type of optimization problems that we consider.

The random modeling of the lives is decisive for the performance of the optimization model. An empirical study of the lives indicated that a non stationary renewal process with Weibull distributed lives was a good model for the recurring maintenance times. This suggests the use of that model directly on the type II engine data. Using physical modeling of the crack size and finer measurements gave a better description of the time to failure/maintenance.

In the example studied, using a discrete distribution for the first life of a stochastic component, resulted in decreased maintenance cost. Here, only a few points of support were necessary for a substantial gain. Further studies are needed to draw conclusions for more realistic situations; however the results seem promising for improving the performance of the optimization model.

Only modeling the first life in the stochastic component as random gives inefficient maintenance decisions, sometimes replacing new components. This calls for developing a finer optimization model treating all recurrent lives in the stochastic components as random. This potentially may blow up the complexity of the optimization, because of the discretization: If one treats the subsequent lives, t_1, \dots, t_k say, as independent random variables the resulting multivariate discretized distribution will have $n = n_1 \cdot \dots \cdot n_k$ points of support, where n_i is the number of support points for the discretized distribution of t_i .

5 Maintenance policies

Currently VAC do not utilize an optimization model for the determination of maintenance schedules. In this section we present the policy that VAC use for this purpose as well as an age replacement policy, which is classical in maintenance planning (cf. Section 1.5). In Section 6 we then evaluate these policies against the optimization model through stochastic simulations.

5.1 A value policy

The methodology currently applied at VAC is a combination of a value policy and manual adjustments.

A tentative replacement schedule for the current maintenance occasion is provided by the following *value policy*. If the remaining life of component $i \in \mathcal{N}$ is \tilde{T}_i the value of the component is $v_i = \tilde{T}_i \cdot c_i / T_i$. Letting d be the fixed cost per maintenance occasion, according to the value policy, a component with $v_i \leq d$ is replaced. If $v_i > d$, component i is *not* replaced.

A problem with this policy is that if component i has a price c_i which is lower than d , then the policy dictates that component i is to be replaced at every maintenance opportunity, regardless of its remaining life. Therefore, the policy is adjusted using a life limit T_{\min} (this value is typically based on customer requirements on the remaining life of the complete engine after maintenance). The adjusted value policy then dictates that if $c_i \leq d$ and $\tilde{T}_i \geq T_{\min}$, then component i is *not* replaced.

The resulting tentative maintenance schedule is then illustrated graphically in an Excel sheet and the user can make manual adjustments in order to provide a cheaper schedule, if possible. At best, this policy may provide schedules that are as good as the ones provided by the optimization model (3), but it would take very great skills to achieve this.

The value policy is developed for safety critical (deterministic) components. On condition (stochastic) components are included by replacing the deterministic life limits with the estimated lives from the conditional expectation.

5.2 An age replacement policy

Age replacement policies are popular in the maintenance literature, e.g., [6]. Each component $i \in \mathcal{N}$ is here given a *life limit* a_i ; if the *age* of component i is higher than a_i then the component is replaced. Finding good values of the life limits is a difficult problem, for which we have implemented the following heuristic procedure. Let $a_i := T_i - \delta$, where $\delta \geq 0$. An optimal value of δ is found by calculating the total maintenance cost using the age replacement policy for the values $\delta = 0, 1, \dots, T$ and picking the value of δ that corresponds to the cheapest maintenance schedule.

Stochastic components are included in this policy analogously as in the value policy.

6 Simulations

We investigate how the three models and methods developed above behave in stochastic situations. For this purpose we create 200 scenarios representing the low pressure turbine’s real behaviour. Table 7 shows its ten components, and indicates which ones are safety critical (SC) or on condition (OC).

Table 7: *Components of the low pressure turbine.*

No.	Component	SC/OC
1	Stator	OC
2	Exhaust frame	SC
3	Roller	SC
4	Seal segment, HPT	OC
5	Case	OC
6	Nozzle segment	OC
7	Shaft, conical	SC
8	Seal, air	SC
9	Disk	OC
10	Blade	OC

The value of the fixed cost d is based on an estimate of the real cost for transport, inspection, administration, etcetera, associated with every maintenance activity regardless of which components are replaced. The value of the time horizon, T , has been set to 1500 flight hours, which is standard procedure at VAC when calculating maintenance prognoses. We do not specify costs explicitly, since this information is classified.

Each SC component has a deterministic life limit. Each OC component is given a Weibull distributed life, which we vary across the simulations.

In order to appreciate the value of performing opportunistic maintenance at all we also compare with the “method” of never replacing an SC component which has not reached its life limit or an OC component which is not considered broken, that is, no opportunistic maintenance is performed. Unless we discretize time this is an unrealistic strategy, since it means that components having only a very small fraction of their lives left still are not replaced during maintenance with the effect that the module must be taken back to the work shop almost the instant it is being used again. Thanks to the discretization made, each time interval consisting of 50 flight hours, it means in our instance that SC components that have a life less than 25 flight hours left will be replaced.

6.1 The deterministic problem

We begin by assuming that all components have deterministic life limits in order to produce a first, deterministic, problem. We hence associate also all OC components with deterministic life limits. Figure 10 shows the results from the four methods. Here, “None” refers to the use of no opportunistic maintenance, as explained in the previous section, “Integer” to the optimization model (3), “Age” to the age replacement policy, and “Value” refers to the value policy currently used at VAC. The notation “Cost” refers to the total maintenance cost over the time interval studied, in relation to the total cost of using no opportunistic maintenance.

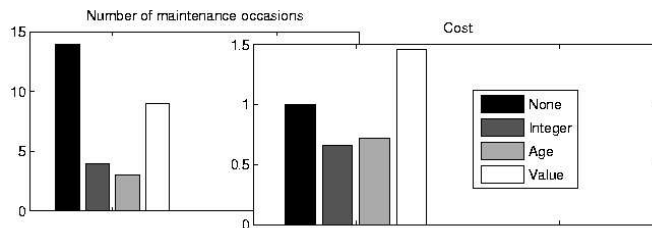


Figure 10: *Number of maintenance occasions and cost for the deterministic problem.*

Figure 11 shows for each component 1–10 how many individuals are replaced for each of the four methods.

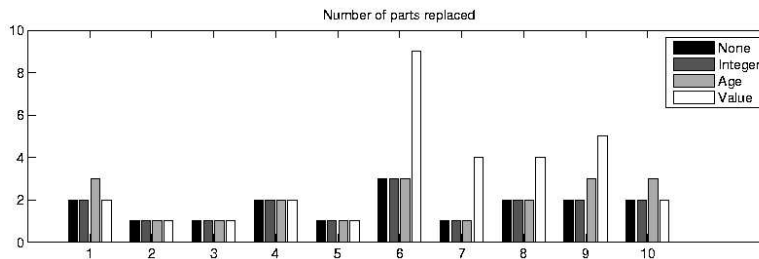


Figure 11: *Number of components replaced for the deterministic problem.*

If no opportunistic maintenance is performed the module is repaired 14 times; each of the opportunistic methodologies reduce this number considerably. The integer model provides the best solution by far in terms of total maintenance costs, whence we see that its use is motivated twofold: both the number of maintenance occasions

and the total cost is reduced considerably. The age replacement policy has a similar behaviour but reduces the number of maintenance occasions even further, however at the cost of replacing components 1, 9, and 10 once too often. (There is no optimal maintenance schedule with less than four maintenance occasions.) Still, the total cost is much lower than when no opportunistic maintenance is performed. The value policy also reduces the number of maintenance occasions, but at the cost of a large number of replacements of components 6, 7, 8, and 9. This is due to the fact that the fixed cost is similar to the cost of each of these components, which has the effect that the value policy dictates that these components are to be replaced (too) often; a close look at the solution shows that 6 replacements of component 6 simply can be stricken. This effect will also be present in our stochastic simulations to follow. Note finally that the total number of replacements of each component is the same in the optimal solution to the integer model and in the case of no opportunistic maintenance, which is also a lower bound on the total number of replacements; the integer model is simply better at grouping these occasions together.

6.2 Stochastic simulations

We next provide results for stochastic simulations with the purpose of learning how opportunistic maintenance fairs when components have stochastic lives. A scenario for an OC component is defined as a sequence of values of (real) lives of the components that may replace an old component at each maintenance opportunity. A scenario for the whole system of OC components is made up by scenarios for each component. In simulations we create 200 such sets of scenarios by drawing deterministic life limits from the respective OC component's life distribution. Following the creation of these 200 scenarios we run the three methods for each scenario and calculate the means of total costs, etcetera. The optimal δ -value obtained in the age replacement policy for the above deterministic problem is utilized in these stochastic simulations. We also apply the method of using no opportunistic maintenance over these scenarios in a similar fashion.

The uncertainty becomes more serious with lower values of the parameter β in the Weibull distribution, and with more stochastic components. Our selection of values of β is based on the knowledge that $2 \leq \beta \leq 6$ for aircraft engine components. In order to investigate the role of the size of β as well as the presence of a larger number of stochastic components, we have run simulations with values 6, 4, and 2 of β , and for each such value we have run tests with a varying number of stochastic components.

First, Figure 12 summarizes the experiments where we have, for components 1, 4, 5, 9, and 10 let the value of β vary among the values 6, 4, and 2, thus gradually

increasing the level of uncertainty.

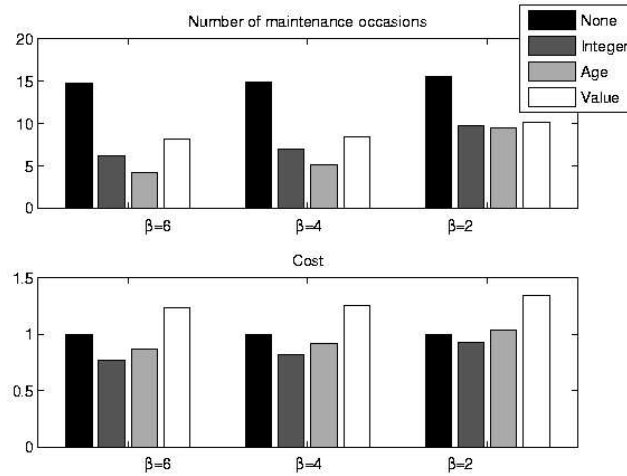


Figure 12: *Number of maintenance occasions and cost. Averages from simulations where $\beta = 6, 4,$ and 2 for components 1, 4, 5, 9, and 10.*

Clearly, maintenance planning becomes more and more difficult as the value of β decreases; however, while the uncertainty is quite substantial in the last example, the integer model still reduces the total cost by 7 % compared to performing no opportunistic maintenance, and for higher values of β the gain is significantly higher still.

Assume finally that components 1 and 4 have $\beta = 2$, components 5 and 6 have $\beta = 4$, and components 9 and 10 have $\beta = 6$. The result of the simulation is shown in Figures 13 and 14.

While the uncertainty is quite substantial in this last example, and therefore the maintenance difficult to plan successfully, the integer model still reduces the total cost by 17 % compared to performing no opportunistic maintenance. The age replacement policy fares less well, and is only marginally better.

In summary, we see that maintenance planning should be performed in an opportunistic manner, even when the uncertainty in the life estimates for the OC components is quite substantial. Using the optimization model always provides a quite large improvement over the current VAC method (no opportunistic maintenance), while the age replacement policy in some cases is even more expensive than the latter. The optimization model provides the best maintenance schedule in each and every case;

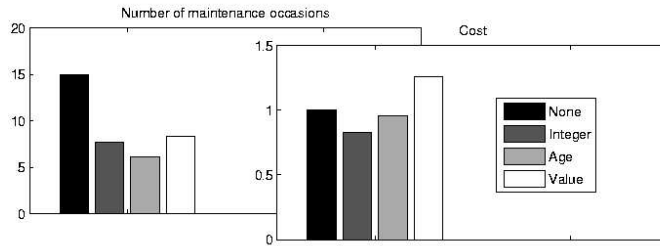


Figure 13: *Number of maintenance occasions and cost. Averages from simulations where $\beta = 2$ for components 1 and 4, $\beta = 4$ for components 5 and 6, and $\beta = 6$ for components 9 and 10.*

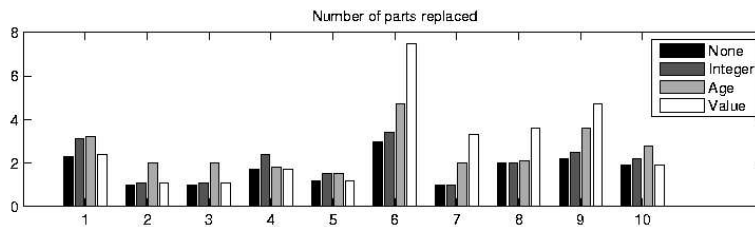


Figure 14: *Number of components replaced. Averages from simulations where $\beta = 2$ for components 1 and 4, $\beta = 4$ for components 5 and 6, and $\beta = 6$ for components 9 and 10.*

the effectiveness of the age replacement policy is however problem dependent—it is not difficult to construct examples when this heuristic provides solutions that are than 50 % more expensive than that provided by the optimization model.

The optimization model also has the clear advantage over all the other ones that it is general, in the sense that more general setting can be relatively easily incorporated. Such extensions could include subsets of the following: additional (side) constraints on the life limits of some (or all) components at the end of the planning period; the presence of a warehouse of cheaper, used spare parts; the consideration of the complete engine, including the associated work costs in disassembling the different modules; and so on. It is not obvious how to extend, for example, the age replacement policy to such additional information.

7 Conclusions

The optimization model described in this paper aims at minimizing the total expected cost during a given time period. The optimization model developed is designed to consider the cost for interrupted production while minimizing the cost of maintenance, in practice meaning that the model will strive to create a maintenance plan with as infrequent maintenance occurrences as possible while maintaining a sound use of replacement parts, new as well as used components.

This is obviously a very useful feature for any organization that needs to operatively schedule and plan the maintenance of any expensive equipment. This type of tool may also be used to create such values that its use can be sold as an additional service product. The described method has been developed and tested for a military aircraft engine, but the potential for use in a commercial context is also encouraging and depends on the kind of agreement between the maintenance provider and the customer. The flight hour agreements mentioned earlier in the text are fairly common within the aero industry.

The usefulness, however, does not end at operative aspects. It also has strategically and tactical uses, for instance, when performing analyses about which components would gain the most on product development, i.e., to get its expected life span prolonged. The engineering work, and cost, required to prolong the life span of a component can be significant. The outlined methodology offers, for example, the opportunity to perform tests in order to select better development projects.

References

- [1] F. ALTENSTEDT, *Aspects on asset liability management via stochastic programming*, PhD thesis, Chalmers University of Technology, 2003.
- [2] P. ANDERSEN, Ø. BORGAN, R. GILL, AND N. KEIDING, *Statistical Models Based on Counting Processes*, Springer-Verlag, New York, 1993.
- [3] N. ANDRÉASSON, *Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems*, licentiate thesis, Department of Mathematics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2004.
- [4] T. W. ARCHIBALD AND R. DEKKER, *Modified block-replacement for multiple-component systems*, IEEE Transactions on Reliability, 45 (1996), pp. 75–83.

- [5] T. AVEN AND R. DEKKER, *A useful framework for optimal replacement models*, Reliability Engineering and System Safety, 58 (1997), pp. 61–67.
- [6] R. BARLOW AND F. PROSCHAN, *Mathematical Theory of Reliability*, John Wiley & Sons, New York, NY, 1965.
- [7] M. BERG AND B. EPSTEIN, *A modified block replacement policy*, Naval Research Logistics Quarterly, 23 (1976), pp. 15–24.
- [8] J. F. BJØRNSTAD, *Predictive likelihood: a review*, Statist. Sci., 5 (1990), pp. 242–265. With comments and a rejoinder by the author.
- [9] D. I. CHO AND M. PARLAR, *A survey of maintenance models for multi-unit systems*, European Journal of Operational Research, 51 (1991), pp. 1–23.
- [10] W. J. COOK, W. H. CUNNINGHAM, W. R. PULLEYBLANK, AND A. SCHRIJVER, *Combinatorial optimization*, John Wiley & Sons, new york, ny ed., 1998.
- [11] R. DEKKER, F. VAN DER DUYN SCHOUTEN, AND R. WILDEMAN, *A review of multi-component maintenance models with economic dependencies*, Mathematical Methods of Operations Research, 45 (1997), pp. 411–435.
- [12] R. DEKKER, R. E. WILDEMAN, AND F. A. V. DUYN-SCHOUTEN, *A review of multi-component maintenance models with economic dependence*, Mathematical Methods of Operations Research, 45 (1997), pp. 411–435.
- [13] B. DICKMAN, S. EPSTEIN, AND Y. WILAMOWSKY, *A mixed integer linear programming formulation for multi-component deterministic opportunistic replacement*, The Journal of the Operational Research Society of India, 28 (1991), pp. 165–175.
- [14] D. B. EDMUNDS, *Modular engine maintenance concept considerations for aircraft turbine engines*, Aircraft Engineering, 50 (1978), pp. 14–17.
- [15] B. FOX, *Age replacement with discounting*, Operations Research, 14 (1966), pp. 533–537.
- [16] J. B. G. FRENK, R. DEKKER, AND M. J. KLEIJN, *A unified treatment of single component replacement models*, tech. rep., Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1996.
- [17] W. J. HOPP AND Y.-L. KUO, *Heuristics for multicomponent joint replacement: Application to aircraft engine maintenance*, Naval Research Logistics, 45 (1998), pp. 435–458.
- [18] ———, *An optimal structured policy for maintenance of partially observable aircraft engine components*, Naval Research Logistics, 45 (1998), pp. 335–352.

- [19] L. M. JR, A. SAXENA, AND G. M. KNAPP, *Statistical-based or condition-based preventive maintenance?*, Journal of Quality in Maintenance Engineering, 1 (1995), pp. 46–59.
- [20] D. KUMAR AND U. WESTBERG, *Maintenance scheduling under age replacement policy using proportional hazards model and TTT-plotting*, European Journal of Operational Research, 99 (1997), pp. 507–515.
- [21] Y. L. KUO, *Scheduling Aircraft Engine Maintenance: Modeling and Optimization*, PhD thesis, Northwestern University, Evanston, IL, 1995.
- [22] P. E. MATHIASSEN, *Prediction functions*, Scand. J. Statist., 6 (1979), pp. 1–21.
- [23] S. MORARU, F. SISAK, AND I. POPESCU, *Overview of preventive maintenance policies for systems, generally, and for digital systems, particularly*, Proceedings of the 6th International Conference on Optimization of Electrical and Electronic Equipments, 3 (1998), pp. 801–806.
- [24] G. L. NEMHAUSER AND L. A. WOLSEY, *Integer and combinatorial optimization*, John Wiley & Sons, New York, NY, 1988.
- [25] Y. PAWITAN, *In all likelihood*, Clarendon press, Oxford, 2001.
- [26] G. C. PFLUG, *Scenario tree generation for multiperiod financial optimization by optimal discretization*, Math. Program., 89 (2001), pp. 251–271. Mathematical programming and finance.
- [27] W. P. PIERSKALLA AND J. A. VOELKER, *A survey of maintenance models: The control and surveillance of deteriorating systems*, Naval Research Logistics Quarterly, 23 (1979), pp. 353–388.
- [28] A. RAN AND S. I. ROSENBLUND, *Age replacement with discounting for a continuous maintenance cost model*, Technometrics, 18 (1976), pp. 459–465.
- [29] Y. SHERIF AND M. SMITH, *Optimal maintenance models for systems subject to failure - a review*, Naval Research Logistics Quarterly, 28 (1981), pp. 47–74.
- [30] J. SVENSSON, *Two statistical models used on aircraft engine data modeling times between repairs. Report*, 2005.
- [31] ———, *Assessment of residual life based on retarding crack growth. Report*, 2006.
- [32] ———, *Discrete approximations of life distributions in optimal replacement. Report*, 2007.
- [33] R. WEBSTER, *Convexity*, Oxford University Press, Oxford, UK, 1994.

A Proofs of the main results of Section 3

Proof of Lemma 11

We derive the result by showing that the constraint matrix of (7) is TU using the characterization in Proposition 8. The inequalities (7) separate over $i \in \mathcal{N}$; therefore it suffices to show that the constraint matrix of the inequality system

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \quad \ell = 1, \dots, T - T_i, \quad (40a)$$

$$-x_{it} \geq -1, \quad t = 1, \dots, T - 1, \quad (40b)$$

is TU for each $i \in \mathcal{N}$.

Let $A^i \in \mathbb{R}^{(T-T_i) \times (T-1)}$ be the constraint matrix defined by the left hand sides of the inequalities (40a), that is,

$$a_{rs}^i = \begin{cases} 1, & s \in \{r, \dots, T_i + r - 1\}, \\ 0, & s \in \{1, \dots, r - 1\} \cup \{T_i + r, \dots, T - 1\}, \end{cases} \quad r \in \{1, \dots, T - T_i\}$$

(here, $\{1, 0\}$ and $\{T, T - 1\}$ should be interpreted as \emptyset). The essential property of the matrix A^i is that the ones appear consecutively in each row, that is, if $a_{r\ell}^i = a_{rk}^i = 1$ and $1 \leq \ell \leq k \leq T - 1$, then $a_{rs}^i = 1$ for all $s \in \{\ell, \dots, k\}$; this property is closed under column deletions. Let $B \in \mathbb{R}^{(T-1) \times (T-1)}$ be the constraint matrix defined by the left hand sides of the inequalities (40b). Then B equals minus the identity matrix, $-I^{T-1}$; if columns are deleted from B , each row will consist of zeros and at most a single -1 . Therefore, it is enough to show that the assumptions in Proposition 8 are satisfied for $J = \{1, \dots, T - 1\}$. Let $J_1 = \{j \in J \mid j \text{ odd}\}$ and $J_2 = J \setminus J_1$. For each $\ell \in \{1, \dots, T - T_i\}$ it holds that

$$\sum_{s \in J_1} a_{\ell s}^i - \sum_{s \in J_2} a_{\ell s}^i = \begin{cases} +a_{\ell, \ell}^i - a_{\ell, \ell+1}^i + \dots - a_{\ell, T_i+\ell-2}^i + a_{\ell, T_i+\ell-1}^i = 1, & \text{if } \ell \text{ odd and } T_i \text{ odd,} \\ +a_{\ell, \ell}^i - a_{\ell, \ell+1}^i + \dots + a_{\ell, T_i+\ell-2}^i - a_{\ell, T_i+\ell-1}^i = 0, & \text{if } \ell \text{ odd and } T_i \text{ even,} \\ -a_{\ell, \ell}^i + a_{\ell, \ell+1}^i - \dots - a_{\ell, T_i+\ell-2}^i + a_{\ell, T_i+\ell-1}^i = 0, & \text{if } \ell \text{ even and } T_i \text{ even,} \\ -a_{\ell, \ell}^i + a_{\ell, \ell+1}^i - \dots + a_{\ell, T_i+\ell-2}^i - a_{\ell, T_i+\ell-1}^i = -1, & \text{if } \ell \text{ even and } T_i \text{ odd,} \end{cases}$$

and for each $\ell \in \{1, \dots, T-1\}$ it holds that

$$\sum_{s \in J_1} b_{\ell s} - \sum_{s \in J_2} b_{\ell s} = \begin{cases} 1, & \text{if } \ell \text{ even,} \\ -1, & \text{if } \ell \text{ odd,} \\ 0, & \text{if column } \ell \text{ deleted.} \end{cases}$$

It follows that the assumptions of Proposition 8(ii) are fulfilled. Hence, the constraint matrix $((A^i)^T, B^T)^T$ of (40) is TU. Since the right-hand sides of (40) are all integral it follows from Proposition 7 that the corresponding polyhedron is integral. \square

Proof of Proposition 12

First note that since $S \subseteq \mathbb{R}^{(N+1)(T-1)}$ it holds that

$$\dim(\text{conv } S) \leq (N+1)(T-1). \quad (41)$$

Then consider the set $Q \subset \mathbb{R}^{N(T-1)}$ consisting of all $x \in \mathbb{R}^{N(T-1)}$ such that

$$\begin{aligned} \sum_{t=\ell}^{T_i+\ell-1} x_{it} &\geq 1, \quad \ell = 1, \dots, T-T_i, \quad i \in \mathcal{N}, \\ x_{it} &\leq 1, \quad t = 1, \dots, T-1, \quad i \in \mathcal{N}. \end{aligned}$$

Let $\bar{x}_{it} = 1$ for all $i \in \mathcal{N}$ and $t = 1, \dots, T$. Then $\bar{x} \in Q$ and since $T_i \geq 2$ for all $i \in \mathcal{N}$, it follows that

$$\sum_{t=\ell}^{T_i+\ell-1} \bar{x}_{it} \geq 2, \quad \ell = 1, \dots, T-T_i, \quad i \in \mathcal{N}.$$

Further, for a given $i \in \mathcal{N}$ and $t \in \{1, \dots, T-1\}$, since $T_i \geq 2$ there exists a vector $\tilde{x} \in Q$ such that $\tilde{x}_{it} = 0$. Hence $\text{rank } A^\# = 0$, where $A^\#$ denotes the rank of the matrix corresponding to the equality subsystem of Q . Proposition 2 then yields that

$$\dim Q = N(T-1).$$

From Lemma 11 it follows that Q is integral and by Proposition 4 we have that Q equals the convex hull of its extreme points. Hence from Proposition 6 it follows that there exists $N(T-1) + 1$ affinely independent integral vectors $y^1, \dots, y^{N(T-1)+1} \in Q$. But this implies that the $N(T-1) + 1$ vectors $q^1, \dots, q^{N(T-1)+1}$ are affinely independent vectors in S , where

$$q^k = \begin{pmatrix} y^k \\ \mathbf{1} \end{pmatrix}, \quad k = 1, \dots, N(T-1) + 1,$$

and $\mathbf{1} \in \mathbb{R}^{T-1}$ is a vector of 1's (corresponding to the z -variables). Now, since $T_i \geq 2$, $i \in \mathcal{N}$, there exist $T - 1$ vectors $q^{N(T-1)+2}, \dots, q^{N(T-1)+T}$ in S , given by

$$q^{N(T-1)+1+k} = \begin{pmatrix} y^{N(T-1)+1+k} \\ \mathbf{1} - e_k \end{pmatrix}, \quad k = 1, \dots, T - 1,$$

where $e_k = (0, \dots, 0, 1, 0, \dots, 0)^T \in \mathbb{R}^{T-1}$ is the k th unit vector. Moreover, it holds that $q^{N(T-1)+1+k} \notin \text{aff} \{q^1, \dots, q^{N(T-1)+k}\}$, $k = 1, \dots, T - 1$, which in turn implies that the vectors $q^1, \dots, q^{N(T-1)+T}$ are affinely independent. Hence, it holds that $\dim(\text{conv } S) \geq N(T-1) + T - 1$. Together with (41) this implies that $\dim(\text{conv } S) = (N + 1)(T - 1)$. \square

Proof of Proposition 14

Since $T_i \geq 2$ for $i \in \mathcal{N}$, $\text{conv } S$ is full-dimensional (Proposition 12). Hence, we can use the uniqueness characterization of the facet description from Proposition 10 to show the assertion.

For each $r \in \mathcal{N}$ and each $\ell \in \{1, \dots, T - T_r\}$, let $\widehat{F}_{r\ell} = \{(x, z) \in \text{conv } S \mid \sum_{t=\ell}^{T_r+\ell-1} x_{rt} = 1\}$. Further, let

$$x_{it}^0 = z_t^0 = 1, \quad i \in \mathcal{N}, \quad t \in \{1, \dots, T - 1\}. \quad (42)$$

Since $T_i \geq 2$ it follows that $(x^0, z^0) \in S \setminus \widehat{F}_{r\ell}$. Then, defining the point $(x^A, z^A) = (x^0, z^0)$ except that $x_{rt}^A = 0$, $t \in \{\ell + 1, \dots, T_r + \ell - 1\}$, it follows that $(x^A, z^A) \in \widehat{F}_{r\ell}$ and hence that $\widehat{F}_{r\ell}$ is a proper face of $\text{conv } S$. Moreover, there exists values on $\lambda \in \mathbb{R}^{N \times (T-1)}$, $\mu \in \mathbb{R}^{T-1}$, and $\rho \in \mathbb{R}$ such that the equation

$$\sum_{t=1}^{T-1} \left(\sum_{i \in \mathcal{N}} \lambda_{it} x_{it} + \mu_t z_t \right) = \rho \quad (43)$$

holds for all $(x, z) \in \widehat{F}_{r\ell}$. We will show that for any value of $\alpha \in \mathbb{R}$ it holds that

$$\lambda_{it} = \begin{cases} \alpha, & \text{if } i = r \text{ and } t \in \{\ell, \dots, T_r + \ell - 1\}, \\ 0, & \text{otherwise,} \end{cases} \quad \begin{cases} \mu_t = 0, & t \in \{1, \dots, T - 1\}, \\ \rho = \alpha. \end{cases}$$

For each $i \in \mathcal{N} \setminus \{r\}$ and each $t \in \{1, \dots, T - 1\}$, let $(x^1, z^1) = (x^A, z^A)$ except that $x_{it}^1 = 0$. It follows that $(x^1, z^1) \in \widehat{F}_{r\ell}$. (x^A, z^A) and (x^1, z^1) , respectively, inserted in (43) then yields that $\lambda_{it} = 0$ for all $i \in \mathcal{N} \setminus \{r\}$ and $t \in \{1, \dots, T - 1\}$.

For each $t \in \{1, \dots, \ell - 1\} \cup \{T_r + \ell + 1, \dots, T - 1\}$, let $(x^2, z^2) = (x^A, z^A)$ except that $x_{rt}^2 = 0$. It follows that $(x^2, z^2) \in \widehat{F}_{r\ell}$. (x^A, z^A) and (x^2, z^2) , respectively, inserted

¹For $\ell \in \{1, T - T_r - 1, T - T_r\}$ the sets $\{1, \dots, \ell - 1\}$ and $\{T_r + \ell + 1, \dots, T - 1\}$, respectively, should be interpreted as \emptyset (and analogously for analogous cases).

in (43) then yields that $\lambda_{rt} = 0$ for all $t \in \{1, \dots, \ell - 1\} \cup \{T_r + \ell + 1, \dots, T - 1\}$. Further, let $(x^B, z^B) = (x^A, z^A)$ except that $x_{r\ell}^B = 0$ and $x_{r, T_r + \ell - 1}^B = 1$, and let $(x^3, z^3) = (x^B, z^B)$ except that $x_{r, T_r + \ell}^3 = 0$. It follows that $(x^3, z^3) \in \widehat{F}_{r\ell}$. (x^B, z^B) and (x^3, z^3) , respectively, inserted in (43) then yields that $\lambda_{r, T_r + \ell} = 0$. The equation (43) can then be rewritten as

$$\sum_{t=1}^{T-1} \mu_t z_t + \sum_{t=\ell}^{T_r + \ell - 1} \lambda_{rt} x_{rt} = \rho \quad (44)$$

For each $t \in \{1, \dots, \ell - 1\} \cup \{\ell + 1, \dots, T_r + \ell - 1\} \cup \{T_r + \ell + 1, \dots, T - 1\}$, let $(x^4, z^4) = (x^A, z^A)$ except that $x_{it}^4 = z_t^4 = 0$, $i \in \mathcal{N}$. It follows that $(x^4, z^4) \in \widehat{F}_{r\ell}$. (x^A, z^A) and (x^4, z^4) , respectively, inserted in (44) then yields that $\mu_t = 0$ for all $t \in \{1, \dots, \ell - 1\} \cup \{\ell + 1, \dots, T_r + \ell - 1\} \cup \{T_r + \ell + 1, \dots, T - 1\}$. Further, for each $t \in \{\ell, T_r + \ell\}$, let $(x^5, z^5) = (x^B, z^B)$ except that $x_{it}^5 = z_t^5 = 0$, $i \in \mathcal{N}$. It follows that $(x^5, z^5) \in \widehat{F}_{r\ell}$. (x^B, z^B) and (x^5, z^5) , respectively, inserted in (44) then yields that $\mu_t = 0$ for all $t \in \{\ell, T_r + \ell\}$. Equation (44) can then be rewritten as

$$\sum_{t=\ell}^{T_r + \ell - 1} \lambda_{rt} x_{rt} = \rho \quad (45)$$

For each $t \in \{\ell + 1, \dots, T_r + \ell - 1\}$, let $(x^6, z^6) = (x^A, z^A)$ except that $x_{r\ell}^6 = 0$ and $x_{rt}^6 = 1$. It follows that $(x^6, z^6) \in \widehat{F}_{r\ell}$. (x^A, z^A) and (x^6, z^6) , respectively, inserted in (45) then yields that $\lambda_{r\ell} = \lambda_{rt}$. Hence, λ_{rt} is constant over $t \in \{\ell, \dots, T_r + \ell - 1\}$ and we define $\lambda_{rt} = \lambda$, $t \in \{\ell, \dots, T_r + \ell - 1\}$. Since $(x^A, z^A) \in \widehat{F}_{r\ell}$ it follows that $\lambda = \rho$. Then, letting $\alpha = \rho$, the equation (45) can now be expressed as

$$\alpha \cdot \sum_{t=\ell}^{T_r + \ell - 1} x_{rt} = \alpha. \quad (46)$$

Proposition 10 then yields that the inequality $\sum_{t=\ell}^{T_r + \ell - 1} x_{rt} \geq 1$ defines a facet of $\text{conv } S$. \square

Proof of Proposition 15

Since $T_i \geq 2$ for $i \in \mathcal{N}$, $\text{conv } S$ is full-dimensional (Proposition 12). Hence, we can use the uniqueness characterization of the facet description from Proposition 10 to show the assertion.

For each $r \in \mathcal{N}$ and each $s \in \{1, \dots, T - 1\}$, let $F_{rs} = \{(x, z) \in \text{conv } S \mid x_{rs} = z_s\}$.

Further, let

$$x_{it}^0 = \begin{cases} 0, & \text{if } i = r \text{ and } t = s, \\ 1, & \text{otherwise,} \end{cases} \quad i \in \mathcal{N}, \quad \text{and} \quad z_t^0 = 1, \quad t \in \{1, \dots, T-1\}.$$

It follows that $(x^0, z^0) \in S \setminus F_{rs}$. Then, defining the point $(x^A, z^A) \in F_{rs}$ as

$$x_{it}^A = z_t^A = 1, \quad i \in \mathcal{N}, \quad t \in \{1, \dots, T-1\}. \quad (47)$$

it follows that F_{rs} is a proper face of $\text{conv } S$. Moreover, there exists values on $\lambda \in \mathbb{R}^{N \times (T-1)}$, $\mu \in \mathbb{R}^{T-1}$, and $\rho \in \mathbb{R}$ such that the equation

$$\sum_{t=1}^{T-1} \left(\sum_{i \in \mathcal{N}} \lambda_{it} x_{it} + \mu_t z_t \right) = \rho \quad (48)$$

holds for all $(x, z) \in F_{rs}$. We will show that for any value of $\alpha \in \mathbb{R}$ it holds that

$$\lambda_{it} = \begin{cases} \alpha, & \text{if } i = r \text{ and } t = s, \\ 0, & \text{otherwise,} \end{cases} \quad \mu_t = \begin{cases} -\alpha, & \text{if } t = s, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad \rho = 0.$$

For each $\ell \in \{1, \dots, T-1\} \setminus \{s\}$, let $(x^1, z^1) = (x^A, z^A)$ except that $x_{r\ell}^1 = 0$. It follows that $(x^1, z^1) \in F_{rs}$. (x^A, z^A) and (x^1, z^1) , respectively, inserted in (48) then yields that $\lambda_{r\ell} = 0$ for all $\ell \in \{1, \dots, T-1\} \setminus \{s\}$.

Similarly, for each $k \in \mathcal{N} \setminus \{r\}$ and each $\ell \in \{1, \dots, T-1\}$, let $(x^2, z^2) = (x^A, z^A)$ except that $x_{k\ell}^2 = 0$. It follows that $(x^2, z^2) \in F_{rs}$. (x^A, z^A) and (x^2, z^2) , respectively, inserted in (48) then yields that $\lambda_{k\ell} = 0$ for all $k \in \mathcal{N} \setminus \{r\}$ and all $\ell \in \{1, \dots, T-1\}$.

For each $\ell \in \{1, \dots, T-1\} \setminus \{s\}$, let $(x^3, z^3) = (x^A, z^A)$ except that $x_{i\ell}^3 = 0$, $i \in \mathcal{N}$, and $z_\ell^3 = 0$. It follows that $(x^3, z^3) \in F_{rs}$. (x^A, z^A) and (x^3, z^3) , respectively, inserted in (48) then yields that $\mu_\ell = 0$ for all $\ell \in \{1, \dots, T-1\} \setminus \{s\}$. Equation (48) can now be rewritten as

$$\lambda_{rs} x_{rs} + \mu_s z_s = \rho. \quad (49)$$

Let $(x^4, z^4) = (x^A, z^A)$ except that $x_{is}^4 = 0$, $i \in \mathcal{N}$, and $z_s^4 = 0$. It follows that $(x^4, z^4) \in F_{rs}$. (x^4, z^4) inserted in (49) yields that $\rho = 0$ and (x^A, z^A) inserted in (49) then yields that $\lambda_{rs} + \mu_s = 0$. Letting $\alpha = \mu_s$ the equation (49) can now be rewritten as

$$\alpha(x_{rs} - z_s) = 0. \quad (50)$$

Hence, Proposition 10 yields that the inequality $x_{rs} \leq z_s$ defines a facet of $\text{conv } S$. \square

Proof of Proposition 16

For each $s \in \{1, \dots, T-1\}$, let $F_s = \{(x, z) \in \text{conv } S \mid z_s = 1\}$. Let $(x^A, z^A) \in F_s$ be defined by (47). Further, let

$$x_{it}^0 = z_t^0 = \begin{cases} 1, & t \in \{1, \dots, T-1\} \setminus \{s\}, \\ 0, & t = s, \end{cases} \quad i \in \mathcal{N}.$$

It follows that $(x^0, z^0) \in S \setminus F_s$ and hence that F_s is a proper face of $\text{conv } S$. Moreover, there exists values on $\lambda \in \mathbb{R}^{N \times (T-1)}$, $\mu \in \mathbb{R}^{T-1}$, and $\rho \in \mathbb{R}$ such that the equation (48) holds for all $(x, z) \in F_s$. We will show that for any value of $\alpha \in \mathbb{R}$ it holds that

$$\begin{aligned} \lambda_{it} &= 0, \quad i \in \mathcal{N}, \quad t \in \{1, \dots, T-1\}, \\ \rho &= \alpha, \end{aligned} \quad \mu_t = \begin{cases} \alpha, & t = s, \\ 0, & t \in \{1, \dots, T-1\} \setminus \{s\}. \end{cases}$$

For each $r \in \mathcal{N}$ and each $\ell \in \{1, \dots, T-1\}$, let $(x^1, z^1) = (x^A, z^A)$ except that $x_{r\ell}^1 = 0$. It follows that $(x^1, z^1) \in F_s$. (x^A, z^A) and (x^1, z^1) , respectively, inserted in (48) then yields that $\lambda_{r\ell} = 0$ for all $r \in \mathcal{N}$ and all $\ell \in \{1, \dots, T-1\}$. Equation (48) can then be rewritten as

$$\sum_{t=1}^{T-1} \mu_t z_t = \rho. \quad (51)$$

For each $\ell \in \{1, \dots, T-1\} \setminus \{s\}$, let

$$x_{it}^2 = z_t^2 = \begin{cases} 1, & t \in \{1, \dots, T-1\} \setminus \{\ell\}, \\ 0, & t = \ell, \end{cases} \quad i \in \mathcal{N}.$$

It follows that $(x^2, z^2) \in F_s$. (x^A, z^A) and (x^2, z^2) , respectively, inserted in (51) yields that $\mu_\ell = 0$ for all $\ell \in \{1, \dots, T-1\} \setminus \{s\}$. Equation (51) can then be rewritten as

$$\mu_s z_s = \rho. \quad (52)$$

Since $z_s = 1$ for all $(x, z) \in F_s$ it follows that $\mu_s = \rho$. Letting $\alpha = \mu_s$, the equation (52) can now be rewritten as $\alpha z_s = \alpha$. Hence, Proposition 10 yields that the inequality $z_s \leq 1$ defines a facet of $\text{conv } S$. \square

Proof of Proposition 17

For each $r \in \mathcal{N}$ such that $T_r \geq 3$ and each $s \in \{1, \dots, T-1\}$ let $\tilde{F}_{rs} = \{(x, z) \in \text{conv } S \mid x_{rs} = 0\}$. Let

$$x_{it}^0 = z_t^0 = 1, \quad i \in \mathcal{N}, \quad t \in \{1, \dots, T-1\}, \quad (53)$$

and let $(x^A, z^A) = (x^0, z^0)$ except that $x_{rs}^A = 0$. Since $(x^0, z^0) \in S \setminus \tilde{F}_{rs}$ and $(x^A, z^A) \in \tilde{F}_{rs}$, it follows that \tilde{F}_{rs} is a proper face of $\text{conv } S$. Moreover, there exists values on $\lambda \in \mathbb{R}^{N \times (T-1)}$, $\mu \in \mathbb{R}^{T-1}$, and $\rho \in \mathbb{R}$ such that the equation (48) holds for all $(x, z) \in \tilde{F}_{rs}$. We will show that for any value of $\alpha \in \mathbb{R}$ it holds that

$$\lambda_{it} = \begin{cases} \alpha, & \text{if } i = r \text{ and } t = s, \\ 0, & \text{otherwise,} \end{cases} \quad i \in \mathcal{N}, \quad \mu_t = 0, \quad \forall t, \quad \text{and} \quad \rho = 0.$$

For each $i \in \mathcal{N}$ and each $t \in \{1, \dots, T-1\}$, let $(x^1, z^1) = (x^A, z^A)$ except that $x_{it}^1 = 0$. Since $T_r \geq 3$, it follows that $(x^1, z^1) \in \tilde{F}_{rs}$. (x^A, z^A) and (x^1, z^1) , respectively, inserted in (48) then yields that $\lambda_{it} = 0$ for all $(i, t) \in \{\mathcal{N} \times \{1, \dots, T-1\}\} \setminus \{(r, s)\}$. Equation (48) can then be rewritten as

$$\lambda_{rs} x_{rs} + \sum_{t=1}^{T-1} \mu_t z_t = \rho. \quad (54)$$

For each $t \in \{1, \dots, T-1\}$, let $(x^2, z^2) = (x^A, z^A)$ except that $x_{it}^2 = z_t^2 = 0$, $i \in \mathcal{N}$. Since $T_r \geq 3$, it follows that $(x^2, z^2) \in \tilde{F}_{rs}$. (x^A, z^A) and (x^2, z^2) , respectively, inserted in (54) then yields that $\mu_t = 0$, $t \in \{1, \dots, T-1\}$. Since $x_{rs} = 0$ for all $(x, z) \in \tilde{F}_{rs}$ we have that $\rho = 0$. Letting $\alpha = \lambda_{rs}$ the equation (54) can be rewritten as $\alpha x_{rs} = 0$ and the proposition follows. \square

Proof of Proposition 20

Aggregating the constraints (11a)–(11d), (11g)–(11h), and (11k) yields the inequality

$$2z_1 + 2x_{12} + 2x_{13} + x_{22} + x_{23} + 2z_4 \geq 3,$$

which is valid for S_{ex} . Multiplying this inequality by $\frac{1}{2}$ results in the valid inequality

$$z_1 + x_{12} + x_{13} + \frac{1}{2}x_{22} + \frac{1}{2}x_{23} + z_4 \geq \frac{3}{2}.$$

Rounding the coefficients in the left-hand side of this inequality up to the nearest integer yields the valid inequality

$$z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4 \geq \frac{3}{2}.$$

We observe that the left-hand side of this inequality will be integral for all points in S_{ex} , so the right-hand side can also be rounded up to the nearest integer, resulting in the inequality (12), which is hence valid for S_{ex} . \square

Proof of Proposition 21

Since $T_i \geq 2$ for $i = 1, 2$, $\text{conv } S_{\text{ex}}$ is full-dimensional (Proposition 12). Hence, we can use the uniqueness characterization of the facet description from Proposition 10 to show the assertion.

Let $F = \{(x, z) \in \text{conv } S_{\text{ex}} \mid z_1 + x_{12} + x_{22} + x_{13} + x_{23} + z_4 = 2\}$. Letting $x_{1t}^0 = x_{2t}^0 = z_t^0 = 1$ for $t = 1, \dots, 4$ it follows that $(x^0, z^0) \in S_{\text{ex}} \setminus F$. Then, defining the point $(x^A, z^A) \in F$ as

$$x_{1t}^A = x_{2t}^A = \begin{cases} 1, & t \in \{1, 4\}, \\ 0, & t \in \{2, 3\}. \end{cases} \quad \text{and} \quad z_t^A = 1, \quad t \in \{1, \dots, 4\},$$

it follows that F is a proper face of $\text{conv } S_{\text{ex}}$. Moreover, there exist values of $\lambda \in \mathbb{R}^{2 \times 4}$, $\mu \in \mathbb{R}^4$, and $\rho \in \mathbb{R}$ such that the equation

$$\sum_{t=1}^4 (\lambda_{1t}x_{1t} + \lambda_{2t}x_{2t} + \mu_t z_t) = \rho \quad (55)$$

holds for all $(x, z) \in F$. We will show that for any value of $\alpha \in \mathbb{R}$ it holds that

$$\lambda_{1t} = \lambda_{2t} = \begin{cases} \alpha, & t \in \{2, 3\}, \\ 0, & t \in \{1, 4\}, \end{cases} \quad \mu_t = \begin{cases} \alpha, & t \in \{1, 4\}, \\ 0, & t \in \{2, 3\}, \end{cases} \quad \text{and} \quad \rho = 2\alpha.$$

For each $s \in \{2, 3\}$, let $(x^1, z^1) = (x^A, z^A)$, except that $z_s^1 = 0$. It follows that $(x^1, z^1) \in F$. (x^A, z^A) and (x^1, z^1) , respectively, inserted in (55) then yields that $\mu_2 = \mu_3 = 0$.

For each $s \in \{1, 4\}$, let $(x^2, z^2) = (x^A, z^A)$, except that $x_{2s}^2 = 0$. It follows that $(x^2, z^2) \in F$. (x^A, z^A) and (x^2, z^2) , respectively, inserted in (55) then yields that $\lambda_{21} = \lambda_{24} = 0$. Equation (55) can then be rewritten as

$$\mu_1 z_1 + \sum_{t=1}^4 \lambda_{1t} x_{1t} + \sum_{t=2}^3 \lambda_{2t} x_{2t} + \mu_4 z_4 = \rho. \quad (56)$$

We define the point $(x^B, z^B) \in F$ as

$$x_{1t}^B = \begin{cases} 1, & t \in \{1, 2\}, \\ 0, & t \in \{3, 4\}, \end{cases} \quad x_{2t}^B = \begin{cases} 1, & t = 1 \\ 0, & t \in \{2, 3, 4\}, \end{cases} \quad z_t^B = \begin{cases} 1, & t \in \{1, 2, 3\}, \\ 0, & t = 4. \end{cases}$$

and let $(x^3, z^3) = (x^B, z^B)$, except that $x_{11}^3 = 0$. It follows that $(x^3, z^3) \in F$. (x^B, z^B) and (x^3, z^3) , respectively, inserted in (56) then yields that $\lambda_{11} = 0$.

We let the point $(x^C, z^C) \in F$ be defined by

$$x_{1t}^C = \begin{cases} 1, & t \in \{2, 4\}, \\ 0, & t \in \{1, 3\}, \end{cases} \quad x_{2t}^C = \begin{cases} 1, & t = 4, \\ 0, & t \in \{1, 2, 3\}, \end{cases} \quad z_t^C = \begin{cases} 1, & t \in \{2, 3, 4\}, \\ 0, & t = 1. \end{cases}$$

and let $(x^4, z^4) = (x^C, z^C)$, except that $x_{14}^4 = 0$. It follows that $(x^4, z^4) \in F$. (x^C, z^C) and (x^4, z^4) , respectively, inserted in (55) then yields that $\lambda_{14} = 0$. Equation (55) can now be expressed as

$$\mu_1 z_1 + \lambda_{12} x_{12} + \lambda_{13} x_{13} + \lambda_{22} x_{22} + \lambda_{23} x_{23} + \mu_4 z_4 = \rho. \quad (57)$$

Let the point $(x^D, z^D) \in F$ be defined by

$$x_{1t}^D = x_{2t}^D = \begin{cases} 1, & t = 2, \\ 0, & t \in \{1, 3, 4\}, \end{cases} \quad z_t^D = \begin{cases} 1, & t \in \{2, 3\}, \\ 0, & t \in \{1, 4\}. \end{cases}$$

and let $(x^5, z^5) = (x^D, z^D)$, except that $x_{22}^5 = 0$ and $x_{23}^5 = 1$. It follows that $(x^5, z^5) \in F$. (x^D, z^D) and (x^5, z^5) , respectively, inserted in (57) then yields that $\lambda_{22} = \lambda_{23}$.

Let $(x^6, z^6) = (x^B, z^B)$, except that $x_{12}^6 = 0$ and $x_{13}^6 = 1$. It follows that $(x^6, z^6) \in F$. (x^B, z^B) and (x^6, z^6) , respectively, inserted in (57) then yields that $\lambda_{12} = \lambda_{13}$.

Since $(x^B, z^B) \in F$ and $(x^C, z^C) \in F$ it follows that $\mu_1 + \lambda_{12} = \mu_4 + \lambda_{12} = \rho$. Hence, $\mu_1 = \mu_4$ and $\lambda_{12} = \rho - \mu_1$. We define $\alpha = \mu_1$ and rewrite equation (57) as

$$\alpha z_1 + (\rho - \alpha)(x_{12} + x_{13}) + \lambda_{22}(x_{22} + x_{23}) + \alpha z_4 = \rho. \quad (58)$$

Since $(x^A, z^A) \in F$ it follows that $\alpha + \alpha = \rho$. Since $(x^D, z^D) \in F$, we have that $\lambda_{22} + \rho - \alpha = \rho$ so that $\lambda_{22} = \alpha$ and rewrite equation (58) as

$$\alpha(z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4) = 2\alpha. \quad (59)$$

Proposition 10 then gives that the inequality (12) defines a facet of $\text{conv } S_{\text{ex}}$. \square

Paper B

Two statistical models used on aircraft engine data modeling times between repairs

Johan Svensson
Chalmers University of Technology
Department of Mathematical Statistics
and
Göteborgs University
Department of Mathematical Statistics
SE-41296 Göteborg, Sweden

Abstract

Military aircraft engines can offer greater operational availability and be more economically maintained through the use of better models that predict times to failure. In this paper, real data consisting of times between repairs of a flame holder in an aircraft engine is used and two models that predict the time to next failure are suggested. The first model is a non-stationary renewal process and the second model is a non-homogeneous Poisson process. A measure to compare the models is defined and with our data the non-stationary renewal process works better. Different repair stations affect the life span of the components but the non-stationary renewal process manages to model this. This model also manages the aging component problem in a effective way. However, in this case no aging is present other than substantial degeneration after the first repair.

Keywords: non-stationary renewal process; non-homogeneous Poisson process; survival;

1 Introduction

Military aircraft engines can offer greater operational availability and be more economically maintained through the use of better models that predict times to failure. Some models exist that use different strategies. Roemer and Ghiocel [12] describe an model that is based on a resistance variable R and a stress variable S . When $R - S < 0$ failure occurs. Tinga and Visser [15] suggest a model that involves a fatigue model, a gas turbine simulation program, fluid dynamic model and a finite

element model. This is combined with a statistical reliability model. Roemer and Ghiocel [13] present a set of tools for health monitoring, diagnostic and prognostic of turbo machinery. One nice tool define two indices and use multivariate process control methods to monitor engines. They also suggest a simple fatigue model for individual components. All of the above models require continuous monitoring of several parameters that are relevant for the life span of the component. They also require that the parameters have historically been stored in such a way that it is possible to make any use of them. Krok and Ashby [10] states that it is hard to develop models based on most of the monitored parameters due high levels of noise and large variation in data.

In other parts of the literature, hundreds of repair and maintenance models can be found. Basic mathematical models and methods can be found in, for instance Høyland and Rausand [5], Klein and Moeschberger [9], Kalbfleisch and Prentice [6] or Andersen et al. [1]. A statistical model that involves physical laws can be found in Yang [19]. Also, pure statistical models can be found in Kaminskiy [7] or Kijima and Masaaki [8]. A survey of some maintenance and reliability models can be found in Weiss, George H. [18].

Based on existing models, this paper will present two models in the class of non-stationary renewal process and non-homogeneous Poisson process to predict failures. The models use historical data for calibration and are suitable for modeling time to failure of stochastic components. A stochastic component is a component that is allowed to operate in the aircraft engine until it breaks. If the component breaks under a flight mission the engine will still work but with reduced performance. A deterministic component is a component that is only allowed to function in the aircraft engine a predetermine specific time. If a deterministic component breaks under a flight mission there is a risk that the engine will stop functioning. In practice this means that the deterministic component is replaced long before its life span is consumed. Methods for calculating the preset times for deterministic components are not discussed here.

There are only two states a component can be in, either functioning or broke. Hence we have a two state system. Recently a lot of work have been made on multistage systems, a survey of multistage system can be found in El-Newehi and Proschan [3]. A detail is considered broken if it fails to comply with a set of fitness rules when it is observed. The flight mission time is short and the details studied here are observed after every flight mission and therefore the time to failure is considered to be the time when the component is observed and fails to comply with the set of fitness rules. In other scenarios, mission times may not be short, or the details are just observed on a few special occasions due to the fact that they are positioned deep inside the engine. The method discussed can still be applied but the derivation of the estimations of the distribution functions may be slightly different.

The data used in this work is gathered from Volvo Aero Corporation in Trollhättan and we are interested in the failure distribution of the flame holder in the

new RM12 engine. We have limited knowledge about this due to the limited number RM12 engines currently in service. We have a large data set containing repair and maintenance times for the same type of details in an older version of the RM8 engine. This data will be of some help in predicting maintenance times for the RM12 engine. This paper will not discuss how to transform knowledge of the RM8 engine to the RM12 engine but will instead discuss two models that predict repair or maintenance times for the RM8 engine.

In the following sections, we present the data material and define two different models. We discuss existing methods to analyze the unknown parameters in the models. An error measure is defined and a comparison between the models is made, using this measure. We modify one of the models so that it can handle different kinds of repairs. Data from the older version of the engine is used. We believe that the models when applied to the new engine design will help us understand the life of the newly designed details.

2 The models

In the literature there are many models, both simple and more complex, that can be applied to predict the time to the next repair. Here we use two models that are in the classes of NSRP (*Non-Stationary Renewal Processes*) respectively NHPP (*Non-Homogeneous Poisson Processes*) to analyse data. We discuss existing methods to analyse the unknown parameters in the models, and close the section with a discussion on measures of goodness of fit.

2.1 The set of data

The data used in this paper originate from a detail called the flame holder which is an essential part in the after burner system. The flame holder consists of two parts called the inner ring and outer ring. The outer ring exists in two versions. We call the inner ring detail 1 and the outer ring detail 2 and 3. We observe the times between repairs. Every time a component fails, a decision is made whether the component should be repaired or scrapped. For every component i we observe a sequence of times between repairs $\{T_1^i, \dots, T_{k_i}^i\}$, $i = 1, \dots, n$, where the last observation possibly is censored, meaning that for the last observation we may have the information that the time to next repair is longer than the time observed. For more information on censoring see Klein and Moeschberger [9]. We model different components as independent, that is $\{T^i\}$ and $\{T^j\}$ are independent vectors if $i \neq j$. We also assume that the censoring process is noninformative, see i.e Kalbfleisch and Prentice [6].

We have more observations of repairs in details 2 and 3 than in detail 1. Doing simple descriptive statistics, we get histogram plots of the number of repairs for each component in Figure 1. The x-axis corresponds to the number of repairs and the y-axis to the number of components that have been repaired x times.

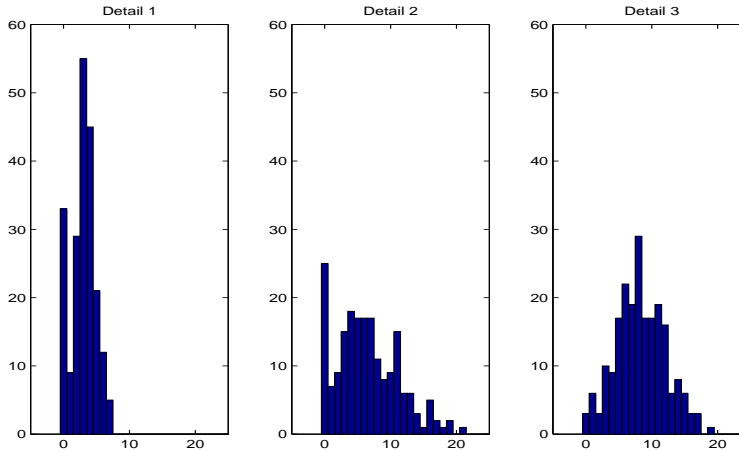


Figure 1: *Histogram on repaired components of each detail. The x-axis corresponds to the number of repairs and the y-axis number of components that have been repaired x times.*

2.2 NSRP-model

A non-stationary renewal process (NSRP) is defined in the following way (cf. e.g. Høyland and Rausand [5]): Let T_n be the time between the $n - 1$ 'th and the n 'th repair, and let $F_n(t) = P(T_n \leq t)$ be the corresponding distribution function. If we assume that the T_n are independently but not equally distributed, the sequence $\{T_n\}_{n \geq 1}$ is called a non-stationary renewal process.

A stationary renewal process is a process that has the same distribution between repairs, whereas the model we use might have different distributions for the successive times between repairs. At this point the question we have to answer is what distribution is suitable to model the time between repairs, that is, what is the distribution of T_n .

We will estimate the distribution between repair times without making any assumptions, that is we will carry out a non-parametric estimation of F_n . The standard approach for censored data goes via the Kaplan-Meier estimator of the survival function $R_n(t) = P(T_n > t)$. This gives us an estimator \hat{F}_n of the cumulative distribution function F_n as follows

$$\hat{F}_n(t) = 1 - \hat{R}_n(t).$$

The Kaplan-Meier estimation of R_n is

$$\hat{R}_n(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{Y(t_i)}\right), \quad (1)$$

where $Y(t_i)$ is the number at risk just before time t_i and d_i is the number of failures at time t_i , see Høyland and Rausand [5] or Klein and Moeschberger [9]. The estimator \hat{F}_n is a step function. Note that when there are no censoring events, \hat{F}_n is the empirical cumulative distribution function.

In order to get an estimate of the probability density function f_n of T_n we can use a kernel estimator

$$\hat{f}_n(t) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t-t_i}{b}\right) \Delta \hat{F}_n(t_i), \quad (2)$$

where K is a kernel function and b is the bandwidth, see Appendix A. The kernel smoothing function used here is

$$K(x) = \frac{15}{16}(1-x^2)^2, \quad x \in [-1, 1], \quad (3)$$

but others may also work well.

A kernel estimate can be informative. If there is no prior knowledge about the distribution, a non-parametric estimate is good as a visual confirmation of the choice of parametric distribution. The advantage of a parametric model is that we automatically extrapolate and get knowledge about the distribution outside the field where we observed data. The parameters may sometimes also have a physical meaning. It is more easy to transfer knowledge from RM8 to RM12 if the model is parametric.

Here the Weibull distribution is used with the following parameterization,

$$F(t) = 1 - e^{-(\frac{t}{\theta})^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0), \quad (4)$$

where θ is the characteristic life and α is the shape parameter. Let n be the number of repairs we are modeling. Then each time between repairs is modeled as a Weibull distributed random variable with its own parameters $\{\alpha_i, \theta_i\}$, so that the parameters to be estimated are $\{\theta_1, \dots, \theta_n\}$ and $\{\alpha_1, \dots, \alpha_n\}$. In a later section we will place restrictions on these parameters.

The parameters were estimated by means of the method of maximum likelihood and the method of least sum of squares, see appendix C and D. To see if the Weibull distribution is a reasonable assumption and get an idea of what estimation method to prefer, density functions from all three estimation methods are plotted in Figure 2.

In Figure 2 kernel functions can be seen as the estimates that are closer to current data. The Maximum likelihood estimate appears closer to the non-parametric estimate than the Least Square estimate. There exist several other formal ways to reach a conclusion, for example the One-sample Test, which with one particular choice of weight function yields the One-sample log-rank test, see Klein and Moeschberger [9]. Another visual way of showing if the choice of parametric distribution function is satisfactory is to plot the function $F^{-1}(\hat{F}(t))$ where \hat{F} is the empirical cumulative

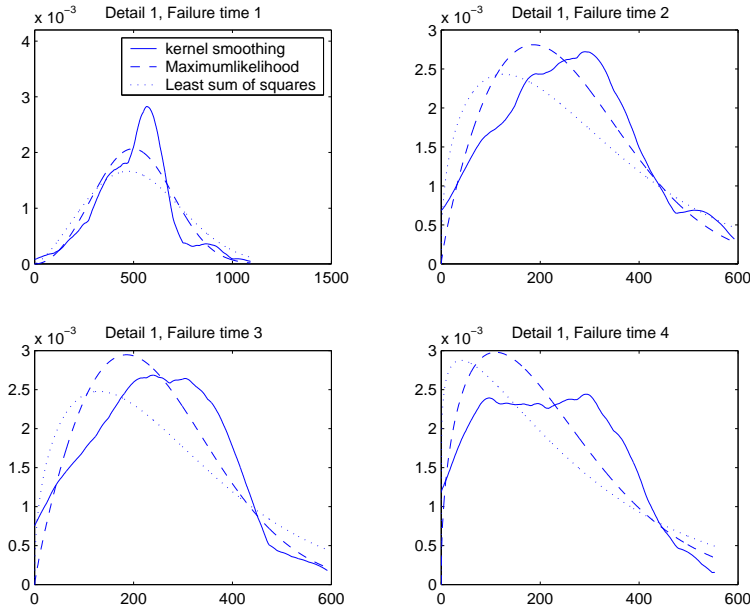


Figure 2: *Maximum likelihood, Least Square and a non-parametric estimation of the density function.*

distribution function from the Kaplan Meier estimation and F is the parametric cumulative distribution function. If F is a good approximation to the data, this should be a straight line. In Figure 3 we see the $F^{-1}(\hat{F}(t))$ function for the four first repair times, with F equal to the Weibull distribution function.

According to this figure Weibull looks like a reasonable approximation for this data set. We see that the Maximum likelihood estimation is closer to the straight line and hence a better approximation. That the maximum likelihood estimator is a better estimator is also indicated in Beretta and Murakami [2]. The choice of the Weibull distribution is not obvious. Other distributions could give a similar approximation of the empirical distribution.

2.3 NHPP-model

Assume that N is a counting process that counts the number of repairs of a component and let the intensity (or hazard) function w be a function of time. If the process has independent and Poisson distributed increments, N is called a Poisson Process. If w is non-constant the process is non-homogeneous. If the number of repairs is Poisson distributed then N is called a non-homogeneous Poisson Process (NHPP).

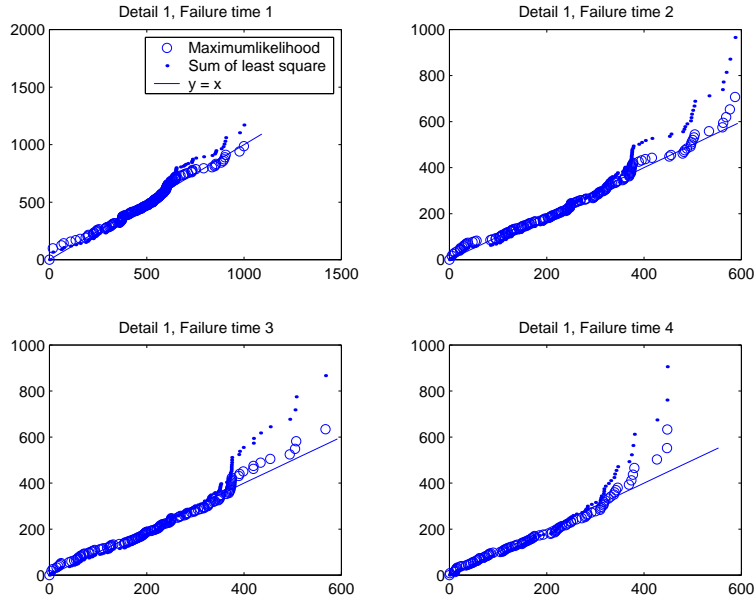


Figure 3: The $F^{-1}(\hat{F}(t))$ function. If the estimation is good the curve should be close to the straight line $y=x$.

If this process is used, there will be dependence between repair times. In reality, modeling with NHPP is the same as assuming minimal repair. That means that when a component is repaired it is repaired to the condition just before the failure occurred. More complicated assumptions can be modeled, see Pham and Wang [11] and Valdez-Flores and Feldman [16]. See Høyland and Rausand [5] for a more theoretical description of Poisson processes and applications.

In this model we can estimate the cumulative intensity directly from the data set. The cumulative intensity is defined as

$$W(t) = \int_0^t w(u) du,$$

and W is estimated by the Nelson-Aalen estimator

$$\hat{W}(t) = \sum_{t_i \leq t} \frac{1}{Y(t_i)}, \quad (5)$$

where $Y(t_i)$ is the number of components at risk just before t_i . The times t_i are all times when failures occur. The distribution of T_1 , the time to the first failure is given

by

$$P(T_1 > t) = P(N(t) = 0) = e^{-W(t)} = e^{-\int_0^t w(u)du},$$

and the distribution of the time T_i to the next failure given a failure at $T_{i-1} = t_{i-1}$ is given by

$$\begin{aligned} P(T_i > t \mid T_{i-1} = t_{i-1}) &= P(N(s_{i-1} + t) - N(s_{i-1}) = 0) = \\ &= e^{-(W(s_{i-1}+t) - W(s_{i-1}))} = e^{-\int_{s_{i-1}}^{s_{i-1}+t} w(u)du}, \end{aligned}$$

where $s_j = \sum_{k=1}^j t_k$ is the absolute time. The function W represents the mean number of repairs for one component until time t . Estimators of these distributions can be obtained by replacing W with \hat{W} , these are related to the Kaplan-Meier estimator but not identical.

Using a kernel smoother we can estimate the intensity function w itself, the quantity is also known as the hazard rate, see Klein and Moeschberger [9]. In Figure 4 estimates of $W(t)$ and $w(t)$ are shown for detail 1, 2 and 3.

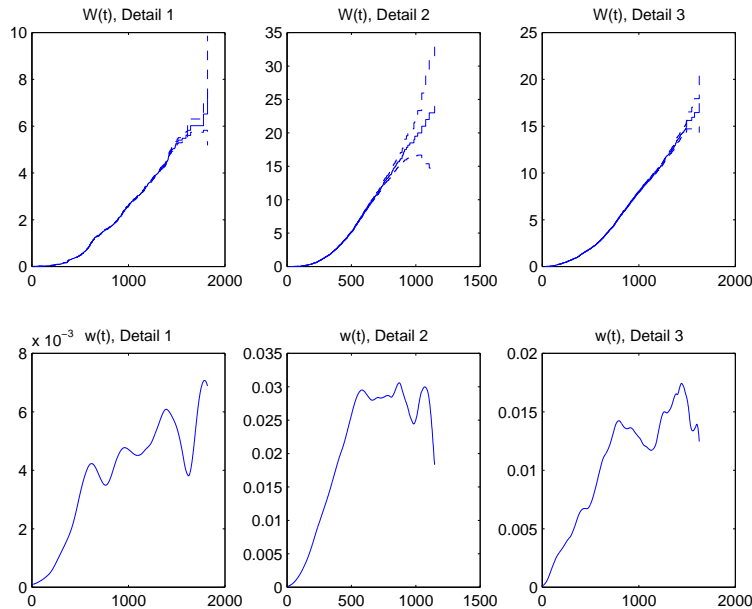


Figure 4: *The Nelson Aalen estimate of W with 90% confidence intervals and kernel estimate of w with kernel function as in (3).*

2.4 Times between failures independent?

When choosing a model it is interesting to know if the times between failures are independent or not. Are the times between repairs related to each other? If the time to first repair comes early is there a greater chance that the component holds longer next time or will the component break early again? Is there any sort of dependency between times of repairs? If the repair times are dependent the NHPP is probably better than the NSRP. In the NSRP the result in the first distribution is independent of the result in the second distribution.

2.5 Which model is the best one?

To decide which model fits data best we must construct some sort of error measurement. There are plenty of measures of fit that are reasonable, e.g.

$$\sup_t | \hat{F}_n(t) - F(t) |, \quad (6)$$

$$\int (\hat{F}_n(t) - F(t))^2 dt, \quad (7)$$

$$\int | \hat{F}_n(t) - F(t) | dt, \quad (8)$$

$$\sum_{i=1}^n (E[T_i] - t_i)^2, \quad (9)$$

$$\sum_{i=1}^n | E[T_i] - t_i |, \quad (10)$$

where $\hat{F}_n(t)$ is the distribution function according to the model, and $F(t)$ is the true distribution function, $E[T_i]$ is the expectation of the model and t_i is the observation. Measure (6) focuses on the greatest difference between the model and the data, (7) and (8) are measures of the overall fit. Measure (9) is a measure of difference in squared mean of the expected outcome and the real outcome, measure (10) is similar to (9) but without square difference penalty. Which measure is the best choice depends on what question we want to answer. In this case we want to use the model to make a prediction of the time when the component fails. As prediction we use the expected failure time according to the model. We want the difference between our prediction and the outcome to be as small as possible. We want a big difference to be more than linear worse than a small difference. This makes it natural to pick (9) as our error measure.

If T_i is the stochastic variable that describes the time to failure, then the error measure m_i for a failure is the quadratic loss function

$$m_i = (E[T_i] - t_i)^2, \quad (11)$$

where t_i is the observed failure time. Finally we take the mean of all components to get the mean square error

$$M = \frac{1}{n} \sum_{i=1}^n m_i.$$

NSRP: It is straightforward to calculate m_i in the NSRP model since T_i is Weibull distributed and $E[T_i]$ then has a known parametric form. With the parameterization as in (4) we have

$$E[T_i] = \theta_i \cdot \Gamma\left(\frac{1}{\alpha_i} + 1\right),$$

where Γ is the gamma function. The expected squared error in the Weibull distribution is

$$M_e = E[m_i] = E[(E[T_i] - T_i)^2] = V[T_i] = \theta^2 \cdot (\Gamma(\frac{2}{\alpha} + 1) - \Gamma^2(\frac{1}{\alpha} + 1)). \quad (12)$$

The reason why the expected error happens to coincide with the variance of the Weibull distribution is a consequence of the fact that we chose the error measure as we did. Table 1 shows the observed and expected squared error for the four first repairs of detail 1,2 and 3. Note that the total error here is the mean of more errors

detail	parameter	failure 1	failure 2	failure 3	failure 4	Total
1	$M * 10^4$	3.28	1.92	1.57	1.61	2.12
1	$M_e * 10^4$	3.41	2.12	1.89	2.50	2.45
2	$M * 10^4$	2.02	0.06	0.06	0.06	0.34
2	$M_e * 10^4$	2.02	0.06	0.06	0.06	0.34
3	$M * 10^4$	7.20	0.22	0.29	0.28	1.05
3	$M_e * 10^4$	7.45	0.21	0.26	0.26	1.08

Table 1: Calculated error in prediction of the NSRP model for detail 1,2 and 3 for the four first time to failure.

than in the four first failure times. It is the mean of all errors to the last failure. As we can see M_e is mostly bigger than M . This may be a consequence of the fact that the true distribution has shorter tails than the Weibull distribution. We can also see the lack of fit in the tail area in Figure 3.

NHPP: In the NHPP the prediction at time t_0 of the time to the next failure T is computed by

$$E[T] = \int_0^{\infty} e^{-(W(t_0+t) - W(t_0))} dt.$$

Because we do not have any estimation of W for large values of t this is impossible to calculate. The standard way of doing this is to stop integration at the largest

observation. Here we present an alternative method of solving this. This method is exact if W is linear and an approximation if W is close to linear. An advantage of this method is that it gives the possibility to check if the NHPP model is appropriate. If we compare the alternative method of calculating with the standard method the total error became 7% bigger with the alternative way for detail 1 but 4% smaller for detail 2 and 3. The idea behind the alternative approach is to transform the NHPP to a HPP with intensity one. This transformation is given by $\hat{t} = W(t)$ where t is the real time and \hat{t} is the transformed time, c.f. Høyland and Rausand [5]. In the HPP the times between errors are independently and exponentially distributed with intensity one. This results in the expected time to next failure being one in the transformed time. We calculate $E[T | T > t_0]$ by means of

1. transform t_0 to \hat{t}_0 ,
2. calculate $E[\hat{T} | \hat{T} > \hat{t}_0] = (\widehat{t_0 + 1})$,
3. transform back $(\widehat{t_0 + 1})$.

We use the alternative way and calculate M for the four first failures. Results are shown in table 2.

detail	parameter	failure 1	failure 2	failure 3	failure 4	Total
1	$M * 10^4$	4.71	1.89	1.65	1.43	2.39
2	$M * 10^4$	2.23	0.42	0.28	0.16	0.43
3	$M * 10^4$	9.77	0.94	0.88	0.72	1.60

Table 2: Calculated error in prediction of the NHPP model for detail 1,2 and 3 in the four first and total time to failure.

If the alternative method is used it is possible to visually control the fact that times between failures are exponentially distributed. Plot the function $F^{-1}(\hat{F}(t))$ where F is the cumulative distribution function of the exponential distribution with intensity one and $\hat{F}(t)$ is the distribution of transformed times between failures. In the data set analysed this is really not the case, especially not for detail 1, as shown in Figure 5. In the HPP, times between failures should be independently distributed if the NHPP is a good model. This can be checked by calculating the correlation between the repair times. The times between repairs are not independently distributed in the data sets examined. This observation indicates that the NHPP-model is not suitable for modeling the behavior of these components.

Comparison A comparison of failure measure between the two models shows that the NSRP model is better modeling this dataset. See Table 3.

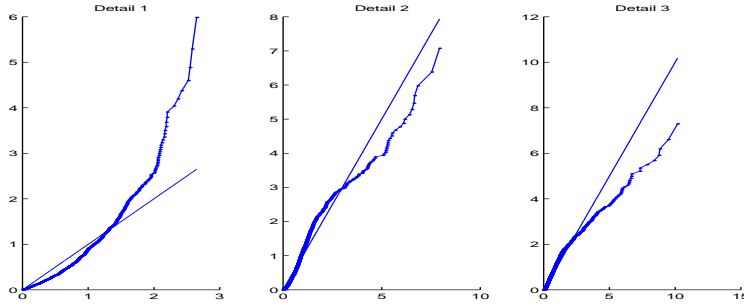


Figure 5: The $F^{-1}(\hat{F}(t))$ function. If the model fit is good the plot follows the straight line $y=x$.

detail	NSRPM * 10 ⁴	NHPPM * 10 ⁴	$\frac{NHPPM}{NSRPM}$
1	2.12	2.39	1.13
2	0.34	0.43	1.26
3	1.05	1.60	1.48

Table 3: Calculated quotient of error measure in the two models

3 Development of the model

Of the two models presented, the NSRP is the better for this type of data. Remember that the main reason for calculating failure times of components for the older engine is that this will help us to understand how similar components of the newer engine will behave. In this section we will see if we can develop the model. A further investigation showed that there are two classes of repairs. We will investigate if these two repairs influence the life of the components differently.

In Figure 2 we can see that the failure distributions seem to be very similar after the first repair. We will use this to refine the model. In Figure 2 we also notice that the time to failure after first repair seemed much longer than the times to the following repairs. This can also be seen in Table 4, if we look at the θ parameter, the characteristic life.

3.1 Different repair stations

There are two different repair stations that repair the details: repair stations within close range of the aircraft A and the main central repair station B. There is reason to believe that the main repair station repairs better than the smaller repair stations.

To investigate if there is a difference between A and B without making any as-

sumptions we estimate the mean time to repair, μ . Let \hat{R} be the Kaplan Meier estimation (1) of the survival function. The mean μ can be estimated by

$$\hat{\mu} = \int_0^\tau \hat{R}(t) dt, \quad (13)$$

where τ is the biggest observed time. The variance of this estimator is

$$V[\hat{\mu}] = \sum_{i=1}^N \left[\int_{t_i}^\tau \hat{R}(t) dt \right]^2 \frac{d_i}{Y_i(Y_i - d_i)}, \quad (14)$$

where N is the number of observations, Y_i the number at risk in time t_i and d_i the number of failures at time t_i . A $100(1 - \alpha)$ confidence interval for μ is expressed by

$$\hat{\mu} \pm Z_{1-\frac{\alpha}{2}} \sqrt{V[\hat{\mu}]}. \quad (15)$$

The process $\hat{R}(t)$ is asymptotically normal distributed so the distribution of $\hat{\mu}$ is also asymptotically normal distributed. Theoretical results can be found in Andersen et al. [1].

We estimate μ and calculate the confidence interval according to (13),(14) and (15) but distinguish between repair *A* and *B*. In Figure 6 this is shown for the first five repairs with 95% confidence interval for the mean. Note that new details are not shown in this figure. (Repair type *B* is displayed as a dot and repair type *A* as a ring. The dots and rings indicate upper and lower confidence bounds. There is also a dot or ring in the middle of the confidence region indicating the point estimation of the mean.) We can see in the figure that there is a difference between the repair *A* and *B*, at least for details 1 and 3. The non-parametric estimate $\hat{\mu}$ is shown in Table 4, p. 15. There it is possible to compare this mean with the mean achieved if the Weibull distribution was assumed. In Table 4 we see that the means are very close, which is an indication that the Weibull distribution is a good approximation.

3.2 Simplify the model

Now when we are aware of the difference in survival depending on where the component has been repaired we model time to next failure with different parameters depending on where the component has been repaired, but we still use different failure distribution after every repair time, that is F_n depends on n . In Table 4 the θ and α parameters are shown for the time to first failure and for the following five times to failure after repair for detail 1,2 and 3. Also μ_w , the mean of Weibull distribution, and μ , the mean of non-parametric distribution, are shown.

We notice that the estimates of θ and α do not change very much between successive repairs if we disregard new components. This makes it natural to suggest a model with the same distribution for T_i , $i \geq 2$. Maximum likelihood estimates of θ and α are shown in Table 5. Note that these values are close to those in Table 4.

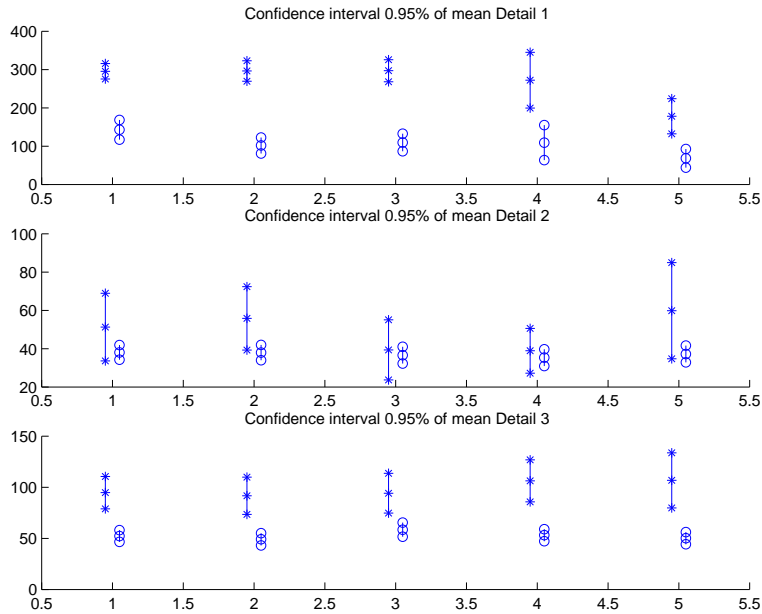


Figure 6: *Confidence interval on expected time to failure after repair number 1 to 5. Dots are repair type B and rings are repair type A.*

When the details in the new engine design begin in service little is known so this may be a good model to start with. Bayesian updating principles may be usable as well, see e.g. Shimi and Tsokos [14], survey or e.g. Michael and Giuntini [4], Weibull example.

3.3 Aging

An interesting question to ask is if the time to next repair decreases with the number of repairs. If we look at the points in Figure 6, do we see a downward trend? A simple test can be made to answer this question. We test the hypotheses that all μ are equal versus that they are not. To carry out this test we must know the numbers of degrees of freedom. This is a complex thing to find out when we have censored data. If we assume that the $\hat{\mu}$ is estimated with many observations the estimated variance of $\hat{\mu}$ is near the true variance. We also assume that the variance of each $\hat{\mu}$ is equal. We can then perform an ordinary χ^2 test, described in appendix B. The results turn out to be that we can only reject the hypothesis in the case detail 3 repair type A. Note that this test only includes the five first repairs. If we only look at the

detail	failure nr	rep type	θ	α	μ_w	μ
1	1	-	554	3.09	496	498
1	2	A	145	1.48	131	131
1	3	A	159	1.46	144	142
1	4	A	109	1.26	101	102
1	5	A	125	1.98	110	110
1	6	A	119	1.37	109	109
1	2	B	359	3.00	321	323
1	3	B	330	3.14	295	295
1	4	B	335	2.83	298	296
1	5	B	331	3.69	299	297
1	6	B	297	3.31	266	272
2	1	-	355	2.36	315	315
2	2	A	45	1.76	40	40
2	3	A	43	1.69	38	38
2	4	A	43	1.76	38	38
2	5	A	41	1.69	37	37
2	6	A	40	1.76	36	35
2	2	B	41	2.20	36	36
2	3	B	52	1.04	51	51
2	4	B	62	1.64	55	56
2	5	B	37	0.84	41	39
2	6	B	44	2.29	39	39
3	1	-	589	1.99	522	526
3	2	A	76	1.77	67	67
3	3	A	58	1.61	52	52
3	4	A	54	1.43	49	49
3	5	A	65	1.49	58	58
3	6	A	59	1.66	53	53
3	2	B	101	1.91	90	90
3	3	B	105	1.47	95	95
3	4	B	98	1.27	91	92
3	5	B	103	1.39	94	94
3	6	B	119	1.68	106	106

Table 4: Parameters θ and α in the Weibull distribution, μ_w the mean in this Weibull distribution and μ the non-parametric estimated mean.

estimated values of $\hat{\mu}$ in Figure 6 we may think that there is a larger difference in μ in detail 2 repair type B, but the confidence interval is much bigger, which indicates that we are less certain of the true value.

detail	repair type	θ	α
1	A	134	1.41
1	B	341	2.98
2	A	41	1.77
2	B	49	1.36
3	A	59	1.51
3	B	107	1.55

Table 5: Parameters θ and α in the Weibull distribution if all T_i , $i \geq 2$ were considered to belong to the same distribution.

Another approach to the aging problem is to assume that the components deteriorate a little every time they are repaired. If they do not, they may in theory be repaired an infinitum number of times and still have the same failure distribution. We suggest the following Weibull model for the T_i , $i \geq 2$

$$F(t) = 1 - e^{-\left(\frac{1}{\theta p^n} t\right)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0, p > 0), \quad (16)$$

where n is the repair number. This means that the expected time to failure after repair number n is

$$E[T_n] = \theta p^n \cdot \Gamma\left(\frac{1}{\alpha} + 1\right),$$

and $p < 1$ thus indicates aging. Maximum likelihood estimates of the parameters (θ, α, p) are shown in Table 6.

detail	repair type	θ	α	p	95% confidence-interval
1	A	164	1.43	0.92	(0.8412 , 1.0146)
1	B	371	3.04	0.95	(0.9205 , 1.0026)
2	A	45.7	1.79	0.98	(0.9666 , 0.9902)
2	B	47.0	1.36	1.01	(0.9205 , 1.0026)
3	A	61.8	1.51	0.99	(0.9794 , 1.0058)
3	B	101.9	1.55	1.01	(0.9846 , 1.0426)

Table 6: Parameters in modified Weibull distribution (16) and a 95% confidence intervals over the parameter p .

Here some \hat{p} are bigger than 1. This is an indication that this may not be a good model. If we make a 95% confidence interval over the true parameter p based on profile likelihood we see that in all cases except one we can not reject that $p = 1$. In the case where we could reject $p = 1$, p is very close to one. This means that if we use $p = 1$ the resulting error is very small. No aging parameter is necessary. If the aging parameter is admitted the model gets one extra parameter that will complicate

the model but add little new information. Profile likelihood is shortly described in appendix E, more can be found in Venzon and Moolgavkar [17].

4 Discussion and conclusion

We have used two models to predict the time between failures on a data set containing failure times of components in an aircraft engine. The model used was a Non-Stationary Renewal Process (NSRP) and a Non-Homogeneous Poisson Process (NHPP). A question arises with model is preferably to use.

In order to understand and measure what model was most suitable to model current data a couple of error measures were considered. The use of the NSRP model was better for this dataset, see Table 3. In the NSRP model non-parametric estimators as Kaplan-Meier (1) and Nelson Aalen (5) were used at an early stage to avoid making any assumptions on parametric distributions. Kernel smoothing (2) was used to analyse times between successive repair times. With the help of the non-parametric estimations a parametric distribution was chosen. Several plots, Figure 2 and 3, showed that the Weibull distribution modeled the data set well enough to make reasonable predictions. For the choice of error measure (11) the expected error in the NSRP model is equal to the variance of the Weibull distribution (12). However, the error was somewhat smaller, which can be a consequence of the fact that the real distribution has smaller tails than the Weibull distribution.

Looking closer at the model, we found that different repair stations had an impact on the times between repairs, see Table 6. We would like to model the time to next failure with different distributions depending on where the component has been repaired. We also noted that the first time to failure was much longer than the following times. However, the following times were from the same distributions. A more advanced model (16) that suggested that the components deteriorated with time, was suggested, but it was rejected by data.

The model we suggest has a distribution to the first failure F_{new} and then two different distributions to the following failures F_A and F_B depending on where the component was repaired. We have 6 parameters to estimate, $\{\theta_{new}, \alpha_{new}, \theta_A, \alpha_A, \theta_B, \alpha_B\}$ see Table 5. We use $\{\theta_{new}, \alpha_{new}\}$ to model the time to next failure if the component is new and $\{\theta_A, \alpha_A\}$ if the component has been repaired at station A, and finally we model with $\{\theta_B, \alpha_B\}$ if the component has been repaired at station B.

The models used data from an older version of an aircraft engine called RM8. A newer engine called RM12 is available but little data exist for that engine. The models in this work can of course be used for RM8 but the main goal was to use them for RM12. That is one of the reasons why a parametric model was chosen since a parametric model may be easier to transfer to the RM12 case than a non-parametric model. However, more work needs to be done in this area.

Acknowledgements

The author wants to thank Dragi Anevski and Thomas Svensson at Fraunhofer Chalmers Center and Jacques de Maré at Chalmers for supervision. Furthermore I own a debt of gratitude to Fredrik Plym and Maud Österman at Volvo Aero Corporation for data support. This work was financed by NFFP.

Appendix

A: Kernel smoothing; Let $H(t)$ be a step function with jumps at the event times $t_1 < t_2 < \dots < t_n$. Let $\Delta H(t_i) = H(t_i) - H(t_{i-1})$ denote the magnitude of the jumps in $H(t_i)$ at time t_i . The kernel smoothed estimator of $h(t)$ is a weighted average of values of $\Delta H(t_i)$ for t_i close to t . Closeness is determined by a bandwidth b so that $t_i \in [t-b, t+b]$ are included in the weighted average. The bandwidth is chosen either to minimise some measure or to give a desired degree of smoothness. Let $K(t)$ be the kernel function that describes how much weight is given to points at a distance from t . Three common kernel functions for $x \in [-1, 1]$ are

$$K(x) = \frac{1}{2}; \quad K(x) = \frac{3}{4}(1-x^2); \quad K(x) = \frac{15}{16}(1-x^2)^2.$$

The estimation is given by

$$\hat{h}(t) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t-t_i}{b}\right) \Delta H(t_i).$$

When $t > t_n - b$ and $t < t_1 + b$ this estimate is biased but can be corrected. More information about kernel smoothing can be found in Klein and Moeschberger [9].

B: χ^2 test; We have 5 estimated means $\{\hat{\mu}_1, \dots, \hat{\mu}_5\}$, each with an estimated variance $\{S_1^2, \dots, S_5^2\}$. We want to test $H_0 : \mu_i$ equal vs $H_1 : \mu_i$ not equal. The number of observations we used to estimate μ_i varies because we use censored data. We assume that μ_i is estimated with a great deal of data and hence S_i^2 is near the true variance. We also assume that all the variances are equal. We calculate

$$S_w = \frac{1}{5} \sum_{i=1}^5 S_i^2; \quad S_b = \frac{1}{4} \sum_{i=1}^5 (\hat{\mu}_i - \bar{\mu})^2; \quad \text{where } \bar{\mu} = \frac{1}{5} \sum_{i=1}^5 \hat{\mu}_i.$$

and

$$\frac{4S_b}{S_w} \sim \chi^2_4.$$

The P-values of this test is shown in Table 7. We can only reject the H_0 hypothesis in the detail 3 repair type A case.

detail	repair type	P-values
1	A	0.245
1	B	0.530
2	A	0.226
2	B	0.599
3	A	0.001
3	B	0.712

Table 7: P-values of the test $H_0 : \mu_i$ equal vs $H_1 : \mu_i$ not equal.

C: Maximum likelihood estimator; We observe $\{t_i, \dots, t_r\}$ as failures and $\{t_{r+1}, \dots, t_n\}$ as censored times. We want to estimate the parameters in (4) by the maximum likelihood method. If the censoring process is random the maximum likelihood estimator is

$$L(\theta, \alpha | t_i) = \prod_{i=1}^r f(\theta, \alpha | t_i) \cdot \prod_{i=r+1}^n R(\theta, \alpha | t_i), \quad (17)$$

and the maximum likelihood estimation, $\hat{\theta}$ and $\hat{\alpha}$ of θ and α is obtained by

$$(\hat{\theta}, \hat{\alpha}) = \underset{\theta, \alpha}{\operatorname{argmax}} L(\theta, \alpha | t_i).$$

D: Least Square Method; To estimate parameters in (4) by the least square method we rewrite F on a form that is linear in the parameters,

$$\log(-\log(1 - F(t))) = \log\left(\frac{t}{\theta}\right)^\alpha = \alpha \log(t) - \alpha \log(\theta). \quad (18)$$

We handle the censored observations by estimate $R(t) = 1 - F(t)$ with Kaplan-Meier estimator (1). We denote

$$\mathbf{Y} = \begin{pmatrix} \log(-\log(\hat{R}(t_1))) \\ \vdots \\ \log(-\log(\hat{R}(t_n))) \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & \log(t_1) \\ \vdots & \vdots \\ 1 & \log(t_r) \end{pmatrix}; \quad \mathbf{P} = \begin{pmatrix} a \\ b \end{pmatrix}$$

where

$$-\alpha \log(\theta) = a; \quad -\alpha = b$$

and consider the equations $\mathbf{Y} = \mathbf{X}\mathbf{P}$. Let $\hat{\mathbf{P}}$ be the estimation of \mathbf{P} and $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{P}}$ the expected failures under $\hat{\mathbf{P}}$. The vector $\hat{\mathbf{P}}$ that minimises $(\mathbf{Y} - \hat{\mathbf{Y}})^T \cdot (\mathbf{Y} - \hat{\mathbf{Y}})$ is called the least square estimation of \mathbf{P} and is calculated by

$$\hat{\mathbf{P}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (19)$$

The parameters θ and α are obtained by

$$\alpha = -b \quad \text{and} \quad \theta = e^{-\frac{a}{b}}$$

One could argue that t_i is the random component and not $\hat{R}(t_i)$. In that case we rewrite (18) to

$$\frac{\log(-\log(R(t)))}{\alpha} + \log(\theta) = \log(t)$$

and define

$$\mathbf{Y} = \begin{pmatrix} \log(t_1) \\ \vdots \\ \log(t_n) \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & \log(-\log(\hat{R}(t_1))) \\ \vdots & \vdots \\ 1 & \log(-\log(\hat{R}(t_n))) \end{pmatrix}; \quad \mathbf{P} = \begin{pmatrix} a \\ b \end{pmatrix}$$

where

$$\log(\theta) = a; \quad \frac{1}{\alpha} = b.$$

We may now minimise $(\mathbf{Y} - \hat{\mathbf{Y}})^T \cdot (\mathbf{Y} - \hat{\mathbf{Y}})$ by (19) and obtain θ and α by

$$\alpha = \frac{1}{b} \quad \text{and} \quad \theta = e^a.$$

We have used the first approach in this report.

E: Profile Likelihood; To make a confidence interval of the parameter p in (16) we first define the profile likelihood

$$\tilde{L}(p) = \max_{\alpha, \theta} L(p, \alpha, \theta),$$

where L is defined in (17) and let

$$\tilde{l}(p) = \log(\tilde{L}(p)),$$

be the log-likelihood function. Let

$$\hat{p} = \operatorname{argmax}_p \tilde{l}(p).$$

Then

$$2(\tilde{l}(\hat{p}) - \tilde{l}(p)) \sim \chi_1^2,$$

and a 95% confidence region is

$$\tilde{R}_c = \left\{ p : \tilde{l}(\hat{p}) - \tilde{l}(p) \leq \frac{\chi_1^2(0.95)}{2} \right\}.$$

References

- [1] P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993.
- [2] S. Beretta and Y. Murakami. Statistical analysis of defects for fatigue strength prediction and quality control of materials. *Fatigue Fract.Engng.Mater.Struct.*, 21:1049–1065, 1998.
- [3] Emad El-Newehi and Frank Proschan. Multistate reliability models: a survey. In *Multivariate analysis, V (Proc. Fifth Internat. Sympos., Univ. Pittsburgh, Pittsburgh, Pa., 1978)*, pages 523–541. North-Holland, Amsterdam, 1980.
- [4] Michael E. Giuntini, Ronald E.; Giuntini. Simulating a weibull posterior using bayes inference. *Annual Reliability and Maintainability Symposium*, pages 48–55, 1993.
- [5] A. Høyland and R. Rausand. *System Reliability Theory*. John Wiley & Sons Inc., 1994.
- [6] J. Kalbfleisch and R. Prentice. *The statistical analysis of failure time data*. John Wiley and sons, New York, 1980.
- [7] V.V. Kaminskiy, M.P.; Krivtsov. G-renewal process as a model for statistical warranty claim prediction. *Annual Reliability and Maintainability Symposium*, pages 276–280, 2000.
- [8] Masaaki Kijima. Generalized renewal processes and general repair models. In *Stochastic models in reliability and maintenance*, pages 145–164. Springer, Berlin, 2002.
- [9] J. Klein and M. Moeschberger. *Survival Analysis*. Springer-Verlag Inc., New York, 1997.
- [10] M. J. Krok and M. J Ashby. Condition-based diagnostic gas path reasoning for gas turbine engines. In *Proceeding of the 2002 IEEE International Conference on Control Applications*, pages 227–234. 2002.
- [11] H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94:425–438, 1996.
- [12] M.J. Roemer and D.M. Ghiocel. Probabilistic approach to the diagnosis of gas turbine engines. Technical report, 53rd Machinery Prvention Technologies (MFPT), Conference, Virginia Beach, VA, April 1999.
- [13] M.J. Roemer and G.J. Kacprzynski. Advanced diagnostics and prognostics for gas turbine engine risk assessment. Technical report, IGTI/ASME Turbo Expo, Munich, Germany, May 2000.

- [14] I. Shimi and C. P. Tsokos. The Bayesian and nonparametric approach to reliability studies: a survey of recent work. In *The theory and applications of reliability, with emphasis on Bayesian and nonparametric methods (Conf., Univ. South Florida, Tampa, Fla., 1975), Vol. I*, pages 5–47. Academic Press, New York, 1977.
- [15] W.P.J. Visser T. Tinga and W.B de Wolf. Integrated lifing analysis for gas turbine components. Technical Report NLR-TP-2000-632, National Aerospace Laboratory NLR, November 2000.
- [16] C. Valdez-Flores and R.M. Fieldman. A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Naval Research Logistics* 36, 419-446, 36:419–446, 1989.
- [17] D.J Venzon and S.H Moolgavkar. A method for computing profile-likelihood-based confidence intervals. *Appl. Statist.*, 37:87–94, 1988.
- [18] George H. Weiss. A survey of some methematical models in the theory of reliability. In *Statistical Theory of Reliability (Proc. Advanced Seminar, Math. Res. Center, U.S. Army, Univ. of Wisconsin, Madison, Wis., 1962)*, pages 3–54. Univ. of Wisconsin Press, Madison, Wis., 1963.
- [19] S.D. Yang, J.N; Manning. A simple second order approximation of stochastic crack growth analysis. *Engineering Fracture Mechanics*, 53:677–686, 1996.

Paper C

Assessment of Residual Life Based on Retarding Crack Growth

Johan Svensson
Chalmers University of Technology
Department of Mathematical Statistics
and
Göteborgs University
Department of Mathematical Statistics
SE-41296 Göteborg, Sweden

Abstract

In laboratory studies the stress strain cycles and other similar parameters are often considered to be known. In real word applications this is not always the case. Still we want to predict the remaining life of components. In a case study the crack growth of the low pressure turbine nozzle is studied. The temperature cycles in the engine result in the nozzle experiencing fatigue and ultimately cracking. A nozzle is considered broken when the largest crack is above a certain level described in a set of fitness rules. The environment in the engine is difficult to model as the temperature profile is unknown due to small air flows through the cracks. Furthermore, the property of the material changes through oxidation. Observations of the nozzle cracks show that the crack growth rate decreases as the crack length increases. In order to model the crack growth and make predictions of the distribution of the remaining time to failure, a new empirical crack model is constructed. The basic components of the model is described by two stochastic variables describing the initiation time of the crack and the crack growth rate. A profile likelihood approach is used to determine both the distribution of the time when the crack reaches a certain size and the distribution of the crack size at a certain time.

Keywords: crack growth; empirical model; predictive profile likelihood, turbine nozzle

1 Introduction

One of many activities at Volvo Aero Corporation (VAC) in Trollhättan is maintaining military aircraft engines. Every time an engine enters the repair bay, a decision on which components in the engine should be replaced must be made. One strategy is to replace only broken components, another is to replace components that are soon to break and hence increase the time to next maintenance occurrence.

In order to optimize the maintenance, there is a need to estimate the remaining life of the components in the engine. When the distribution of the remaining life is known we can use an optimization model, Andréasson [1], or some other policy, cf. Wang [15], to minimize the expected cost of maintaining the engine.

A component is considered broken when it fails to comply with a set of fitness rules. When a broken component is observed it must be replaced. The fitness rules or the failure criteria are measured in quantitative variables, e.g. crack length. When an engine is examined, observations of these variables become available and can be used to estimate the life of a component.

We consider the low pressure turbine nozzle component in a military aircraft engine. From experience we know that cracks are the most common cause of failure of the low pressure turbine nozzle component. Old data, from similar components, tells us that 96-98% of failures are due to cracks, and it is therefore natural to use a model for crack growth.

The historical data we have is limited and we do not know the exact nature of the crack growth. The component experiences thermodynamic cycles which means that the crack growth is temperature driven. If we know what missions the plane is going to experience in the future, we can use those mission profiles and a thermodynamic model to try to estimate the loads. In this case we are not sure about future missions and we do not have a complete understanding of the thermodynamic environment. When a crack starts growing we get airflow through the crack complicating the heat profile even more. There may also be oxidation in the crack further complicating the physics behind the crack growth.

We only have measurements of the crack on a few occasions, but the observations show that the crack growth rate decreases as the crack length increases. Even without the difficulties above, crack growth is stochastic in nature, cf. Virkler et al. [14], Bolotin [5], Yang [16]. If we want to predict the crack growth and control uncertainties we need a stochastic model.

In our case we do not know the mission profiles but from experience and by examining logged thermodynamic cycles for different engines we know that the load sequence is fairly similar over longer time periods, that is more than a few missions. Since we only have limited knowledge about the physics involved we decide to construct an empirical crack growth model that fits our observations. In the model we assume that load cycles are correlated with the flight time. The modeling and analysis done in this paper rest on the assumption that components that experience

similar loads will have similar crack growth behavior.

From observations we conclude that the crack growth rate decreases as the crack increases. Retarding crack growth is uncommon but arises from that the stress in the crack tip decreases when the crack growth longer often due to geometrical reasons. D. B. Garcia et al. [7] study decreasing crack growth in an aircraft nose landing gear drag brace fitting.

This paper is divided into two parts: one modeling part and one case study part. In the modeling part a stochastic crack growth model is presented. We also derive the formulas used to make predictions about the life distributions of the components and the crack length distribution at a fixed time. The second part is a case study where we use data from Volvo to learn how the model reacts.

2 Notation

We use the following Notations:

n	number of cracks
n_i	number of observations of crack i
t	time
a_{max}	a fixed crack length
T	stochastic time when crack reaches length a_{max}
A	stochastic crack length of crack at fixed time t
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$\Phi(x)$	$N(0, 1)$ cumulative distribution function at point x
$F_{c,s}$	distribution function of stochastic variables C and S
x_{ij}	the crack length observation of crack i at time t_{ij}
\mathbf{x}_i	n_i observations of crack i , $\{x_{i,1}, \dots, x_{i,n_i}\}$
$\underline{\mathbf{x}}$	observations $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ of all n cracks
$\underline{\mathbf{x}}_{-i}$	observations $\{\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n\}$ excluding crack i

3 Modeling part

In the modeling part we define the crack growth model and derive estimates of the parameters in the model. We also derive the distribution of time to failure and the distribution of the crack length at a fixed time. First we consider all model parameters known and make a probability model, then we consider the noise we have in the parameter estimation and make a statistical model that takes the parameter uncertainty into account. We use a predictive profile likelihood approach, cf. Pawitan [12] and Björnstad [4].

3.1 Crack model

The model we use for crack growth is

$$a(t) = \begin{cases} a_0, & \text{if } t < S, \\ a_0 + C(t - S)^b, & \text{if } t \geq S, \end{cases} \quad (1)$$

where $a(t)$ is the crack length at time t , C and S are two stochastic variables and $b > 0$ is a parameter that describes how the crack is growing over time. For each crack we have, one outcome of S and C while b is constant for all cracks. The variable S is the time when the crack has grown beyond length a_0 . We are not interested in modeling crack growth below length a_0 . The reason for this is that it may be hard to detect small cracks and also that the crack growth curve we suggest here may not be applicable to small cracks. The C variable is related to how fast the crack is growing when it passes the length a_0 . We assume that both C and S are positive continuous stochastic variables with a bivariate parametric distribution $F_{C,S}(c, s; \psi)$ with parameter vector ψ .

For each crack $i = 1, \dots, k$ we have n_i observations $x_{ij} = x_{ij}(t_{ij})$ at times $t_{i1} < \dots < t_{in_i}$. The observations are assumed to follow the (conditional) model

$$X_{ij}(t_{ij}) = \begin{cases} a_0 & \text{if } t_{ij} < S_i, \\ \max(a_0, a(t_{ij}) + \varepsilon_{ij}) & \text{if } t_{ij} \geq S_i, \end{cases}$$

with ε_{ij} a sequence of independent $N(0, \sigma_\varepsilon^2)$ distributed random variables. Thus the distribution of $X_{ij}(t_{ij})$ is conditionally on (C_i, S_i) a mixture of a discrete and continuous distribution, the discrete random variable having a point mass at a_0 .

The observation model implies that we never detect a crack below length a_0 but that we sometimes fail to observe a crack above length a_0 . We gather all parameters in the model in $\theta = (\psi, \sigma_\varepsilon^2, b)$.

3.2 Probability model: Time distribution until failure

For crack i we want to find the the distribution of the time T until the crack length $a(t)$ reaches a specified maximum length $a_{max} > a_0$,

$$T = \inf\{t : a(t) \geq a_{max}\}.$$

Assume that we are given previous observations \mathbf{x}_i of crack i , and assume the parameters θ of the model are known to us.

Using model (1), we find

$$\begin{aligned} P(T \leq t \mid \mathbf{X}_i = \mathbf{x}_i) &= P(a(t) \geq a_{max} \mid \mathbf{X}_i = \mathbf{x}_i) \\ &= \int \int I\{a(t) \geq a_{max}\} f_{(C,S) \mid \mathbf{x}_i}(c, s \mid \mathbf{x}_i; \theta) dc ds \\ &= \int \int I\{s \leq t - (\frac{a_{max} - a_0}{c})^{\frac{1}{b}}\} f_{(C,S) \mid \mathbf{x}_i}(c, s \mid \mathbf{x}_i; \theta) dc ds. \end{aligned} \quad (2)$$

When differentiating inside the integral in (2) we obtain the conditional density

$$f_{T|\mathbf{x}_i}(t|\mathbf{x}_i; \theta) = \int f_{(C,S)|\mathbf{x}_i}(c, t - (\frac{amax - a_0}{c})^{\frac{1}{\tau}} | \mathbf{x}_i; \theta) dc. \quad (3)$$

Differentiating is possible since $f_{(C,S)|\mathbf{x}_i}$ is a bivariate density function, cf. e.g. Apostol [2] p. 283.

Using Bayes' formula we rewrite the integrand

$$f_{(C,S)|\mathbf{x}_i}(c, s | \mathbf{x}_i; \theta) = \frac{f_{\mathbf{x}_i|C,S}(\mathbf{x}_i | c, s; b, \sigma_\varepsilon^2) f_{C,S}(c, s; \psi)}{f_{\mathbf{x}_i}(\mathbf{x}_i; \theta)}. \quad (4)$$

In order to derive the joint density $f_{\mathbf{X}_i}(\mathbf{x}_i; \theta)$ of the crack lengths for crack i we introduce $t_{i0} = 0$ and $t_{i, n_i+1} = \infty$ and let \leq denote the partial order on \mathbb{R}^n .

(i) Assume first that $\mathbf{x}_i \geq a_0$. Note that if $S_i \in (t_{i, j-1}, t_{ij})$ for $j \leq n_i$, then

$$\begin{aligned} \{\mathbf{X}_i \leq \mathbf{x}_i\} &= \cap_{k=1}^{n_i} \{X_{ik} \leq x_{ik}\} \\ &= \cap_{k=1}^{n_i} (\{t_{ik} < S_i\} \cup \{t_{ik} \geq S_i\}) \{X_{ik} \leq x_{ik}\} \\ &= \cap_{k=1}^{n_i} (\{t_{ik} < S_i\} \{a_0 \leq x_{ik}\} \\ &\quad \cup \{t_{ik} \geq S_i\} \{\max(a_0, a(t_{ik}) + \varepsilon_{ik}) \leq x_{ik}\}) \\ &= \cap_{k=j}^{n_i} \{a(t_{ik}) + \varepsilon_{ik} \leq x_{ik}\}, \end{aligned} \quad (5)$$

where the last equality follows since $\{a_0 \leq x_{ik}\} = \Omega$, $\{t_{ik} < S_i\} = \emptyset$ and $\{t_{ik} \geq S_i\} = \Omega$. If $S_i > t_{i, n_i}$ we have instead

$$\{\mathbf{X}_i \leq \mathbf{x}_i\} = \cap_{k=1}^{n_i} \{X_{ik} \leq x_{ik}\} = \cap_{k=1}^{n_i} \{a_0 \leq x_{ik}\} = \Omega \quad (6)$$

since we assumed $\mathbf{x}_i \geq a_0$.

(ii) Assume next that $x_{i, j} < a_0$ for all j in some nonempty subset $J_i = \{j\}$ of $\{1, \dots, n_i\}$ and that $x_{i, k} \geq a_0$ for $k \in J_i^c = \{1, \dots, n_i\} - J_i$. Then

$$\{\mathbf{X}_i \leq \mathbf{x}_i\} = \emptyset$$

since the observation x_{ik} are bounded below by a_0 .

Thus the sets $\{\mathbf{X}_i \leq \mathbf{x}_i\}$ are nonempty only when $\mathbf{x}_i \geq a_0$. This implies that the distribution for $\mathbf{X}_i = (X_{i1}, \dots, X_{i, n_i})$ is nonzero only when $\mathbf{x}_i \geq a_0$, which we assume

next. Then from (5) and (6) we obtain

$$\begin{aligned}
P(\mathbf{X}_i \leq \mathbf{x}_i) &= \sum_{j=1}^{n_i+1} P\left(t_{i,j-1} < S_i \leq t_{ij}, \mathbf{X}_i \leq \mathbf{x}_i\right) \\
&= \sum_{j=1}^{n_i} P\left(\{t_{i,j-1} < S_i \leq t_{ij}\} \cap \bigcap_{k=j}^{n_i} \{a(t_{ik}) + \varepsilon \leq x_{ik}\}\right) \\
&\quad + P\left(t_{in_i} < S_i \leq t_{i,n_i+1}\right) \\
&= \sum_{j=1}^{n_i} \int_{s=t_{i,j-1}}^{t_{ij}} \int \prod_{k=j}^{n_i} \Phi\left(\frac{x_{ik} - a(t_{ik})}{\sigma_\varepsilon}\right) f_{C,S}(c,s) dc ds, \\
&\quad + 1 - P\left(S_i \leq t_{i,n_i}\right) \\
&=: F^{(1)}(\mathbf{x}_i; \theta), \tag{7}
\end{aligned}$$

where the next to last equality follows since x_{ij} has, conditionally on (C_i, S_i) and on the interval $[S_i, \infty)$, a Gaussian distribution. Thus

$$F_{\mathbf{X}_i}(\mathbf{x}_i; \theta) = F^{(1)}(\mathbf{x}_i; \theta) I\{\mathbf{x}_i \geq \mathbf{a}_0\}, \tag{8}$$

where $I\{\mathbf{x}_i \geq \mathbf{a}_0\} = 1\{x_{i1} \geq a_0\} \cdot \dots \cdot 1\{x_{in_i} \geq a_0\}$. Also introduce the notation $\delta_{\mathbf{a}_0}(\mathbf{x}_i) = \delta_{a_0}(x_{i1}) \cdot \dots \cdot \delta_{a_0}(x_{in_i})$, and note that this is a well-defined distribution function on \mathbb{R}^{n_i} .

Introduce multi-indices α as tuples of elements from $N_{n_i} = \{1, \dots, n_i\}$, e.g. $\alpha = (1, 4)$. Also let α' be the complementary indexes from N_{n_i} , so that α, α' are both tuples elements from N_{n_i} whose union is all of N_{n_i} . Let ∂^α denote partial derivation with respect to the elements in α , e.g. $\partial^{(1,4)}u(x) = \frac{\partial}{\partial_1} \frac{\partial}{\partial_4} u(x)$, and let $(\prod_{i=1}^n u_i(x_i))^\alpha$ denote the partial product over α , e.g. $(\prod_{i=1}^n u_i(x_i))^{(1,4)} = u_1(x_1)u_4(x_4)$, for u and $u_i, i = 1, \dots, n$, arbitrary functions.

The density function is then, via (8), given by

$$f_{\mathbf{X}_i}(\mathbf{x}_i; \theta) = \sum_{\alpha: |\alpha|+|\alpha'|=n_i} \partial^\alpha F^{(1)}(\mathbf{x}_i; \theta) I\{\mathbf{x}_i \geq \mathbf{a}_0\}^\alpha \delta_{\mathbf{a}_0}(\mathbf{x}_i)^{\alpha'}. \tag{9}$$

Thus \mathbf{X}_i is an n_i -variate r.v., that is a mixture of 2^{n_i} r.v.'s, say $\Xi_{i1}, \dots, \Xi_{i2^{n_i}}$ that live either on the interval $[a_0, \infty)^{\times n_i}$, in which case the corresponding r.v. Ξ_{ij} is a continuous random variable, or live and on the edges of the interval $[a_0, \infty)^{\times n_i}$; in particular one of the Ξ_{ij} is a discrete random variable with mass $F^{(1)}(\mathbf{a}_0; \theta)$ in \mathbf{a}_0 .

Using (7) we see that a typical term in (9) is of the form

$$\begin{aligned}
& \partial^\alpha F^{(1)}(\mathbf{x}_i; \theta) I\{\mathbf{x}_{ij} \geq a_0\}^\alpha \delta_{a_0}(\mathbf{x}_i)^{\alpha'} \\
= & I\{\mathbf{x}_{ij} \geq a_0\}^\alpha \delta_{a_0}(\mathbf{x}_i)^{\alpha'} \left(\sum_{j=1}^n \int_{s=t_{i,j-1}}^{t_{i,j}} \int \left[\prod_{k=j}^{n_i} \Phi\left(\frac{-c(t_{ij} - s)^b}{\sigma_\varepsilon}\right) \right]^{\alpha'} \right. \\
& \left. \left[\prod_{k=j}^{n_i} \frac{1}{\sigma_\varepsilon} \phi\left(\frac{x_k - a_0 - c(t_{ij} - s)^b}{\sigma_\varepsilon}\right) f_{C,S}(c, s) dc ds \right]^\alpha \right. \\
& \left. \prod_{k=1}^{j-1} (1 - I\{k \in \alpha\}) + [1 - P(S \leq t_{i,n_i})] 1\{\alpha = \emptyset\} \right) \quad (10)
\end{aligned}$$

where $\prod_{k=1}^{j-1} (1 - I\{k \in \alpha\})$ occurs due to the fact that for some j the terms in $F^{(1)}$ disappear in the differentiation.

The numerator $f_{\mathbf{x}_i|C,S}(\mathbf{x}_i | c, s; b, \sigma_\varepsilon^2)$ in (4) is obtained from (i) and (ii) similar to the derivation of (7) as

$$\begin{aligned}
P(\mathbf{X}_i \leq \mathbf{x}_i | C_i = c_i, S_i = s_i) &= \sum_{j=1}^{n_i} \{t_{i,j-1} < s_i < t_{ij}\} \\
& \prod_{k=j}^{n_i} \Phi\left(\frac{x_{ij} - a_0 - c_i(s_i - t_{ij})^b}{\sigma_\varepsilon}\right) + I\{s_i > t_{i,n_i}\} \\
& =: F^{(2)}(\mathbf{x}_i | c_i, s_i; b, \sigma_\varepsilon^2). \quad (11)
\end{aligned}$$

Now

$$\begin{aligned}
& f_{\mathbf{x}_i|C,S}(\mathbf{x}_i | c_i, s_i; b, \sigma_\varepsilon^2) = \\
& \sum_{\alpha: |\alpha| + |\alpha'| = n} \partial^\alpha F^{(2)}(\mathbf{x}_i | c_i, s_i; b, \sigma_\varepsilon^2) I\{\mathbf{x}_i \geq a_0\}^\alpha \delta_{a_0}(\mathbf{x}_i)^{\alpha'},
\end{aligned}$$

where a typical term looks like this

$$\begin{aligned}
& \partial^\alpha F^{(2)}(\mathbf{x}_i; \theta) I\{\mathbf{x}_i \geq a_0\}^\alpha \delta_{a_0}(\mathbf{x}_i)^{\alpha'} \\
= & I\{\mathbf{x}_i \geq a_0\}^\alpha \delta_{a_0}(\mathbf{x}_i)^{\alpha'} \left(\sum_{j=1}^n I\{t_{i,j-1} < s_i \leq t_{ij}\} \right) \quad (12)
\end{aligned}$$

$$\begin{aligned}
& \left[\prod_{k=j}^{n_i} \Phi\left(\frac{-c_i(t_{ij} - s_i)^b}{\sigma_\varepsilon}\right) \right]^{\alpha'} \left[\prod_{k=j}^{n_i} \frac{1}{\sigma_\varepsilon} \phi\left(\frac{x_k - a_0 - c_i(t_{ij} - s_i)^b}{\sigma_\varepsilon}\right) \right]^\alpha \\
& \prod_{k=1}^{j-1} (1 - I\{k \in \alpha\}) + I\{s_i > t_{i,n_i}\} 1\{\alpha = \emptyset\} \quad (13)
\end{aligned}$$

Having derived all factors in (4), we now have an expression for the distribution of the time to failure.

3.3 Probability model: Crack distribution at a fixed time

Let A be the crack length of crack i at a fixed time t . We want to find the distribution of A , given previous observations \mathbf{x}_i of crack i . From model (1) it follows that $A \geq a_0$. We start by finding

$$f_{A|\mathbf{X}_i}(a, | \mathbf{x}_i; \theta) = \frac{f_{\mathbf{X}_i|A}(\mathbf{x}_i | a; \theta) f_A(a; \theta)}{f_{\mathbf{X}_i}(\mathbf{x}_i; \theta)}. \quad (14)$$

We rewrite the second factor in the numerator in (14),

$$\begin{aligned} f_A(a; \theta) &= I\{a > a_0\} \frac{d}{da} P(A \leq a | \theta) + \delta_{a_0}(a) P(A = a | \theta) \\ &= I\{a > a_0\} \frac{d}{da} \int \int I\{a(t) \leq a\} f_{S,C}(s, c; \theta) ds dc \\ &\quad + \delta_{a_0}(a) P(S_i > t) \\ &= I\{a > a_0\} \frac{d}{da} \int \int_{s=0}^t I\{a_0 + c(t-s)^b \leq a\} f_{S,C}(s, c; \theta) ds dc \\ &\quad + \delta_{a_0}(a) (1 - P(S \leq t)) \\ &= I\{a > a_0\} \frac{d}{da} \int \int_{s=0}^t I\{c \leq \frac{a - a_0}{(t-s)^b}\} f_{C,S}(c, s; \theta) ds dc \\ &\quad + \delta_{a_0}(a) (1 - F_S(t; \theta)) \end{aligned} \quad (15)$$

$$= I\{a > a_0\} \int_{s=0}^t \frac{f_{C,S}(\frac{a-a_0}{(t-s)^b}, s)}{(t-s)^b} ds + \delta_{a_0}(a) (1 - F_S(t; \theta)) \quad (16)$$

where we assume that we can differentiate inside the integral. Differentiation is possible if $f_{C;S}$ provided that we can find a nonnegative function G such that

$$\left| \frac{f_{C,S}(\frac{a-a_0}{(t-s)^b}, s)}{(t-s)^b} \right| \leq G(s), \quad \text{and} \quad \int_{s=0}^t G(s) ds < \infty,$$

cf. Apostol [2] p. 283. If this is not the case we use (15) instead of (16).

The function $f_{\mathbf{X}_i}(\mathbf{x}_i; \theta)$ in the denominator of (14) is given in (9). Now we know every function in (14) except $f_{\mathbf{X}_i|A}(\mathbf{x}_i | a; \theta)$. We start by deriving the distribution $P(\mathbf{X}_i \leq \mathbf{x}_i | a(t) = \eta)$ by splitting into the two cases $\eta = a_0$ and $\eta > a_0$. In the

derivation below, assume that $\mathbf{x}_i \geq a_0$. When $\eta = a_0$,

$$\begin{aligned}
P(\mathbf{X}_i \leq \mathbf{x}_i \mid a(t) = \eta) &= \sum_{j=1}^{n_i+1} I\{t_{i,j-1} \leq t < t_{ij}\} \\
&\quad \sum_{m=1}^{n_i+1} P(\{t_{i,m-1} \leq S_i < t_{i,m}\} \{\mathbf{X}_i \leq \mathbf{x}_i\} \mid \{t \leq S_i\}) \\
&= \sum_{j=1}^{n_i+1} I\{t_{i,j-1} \leq t < t_{ij}\} \\
&\quad \sum_{m=1}^{n_i+1} \frac{P(\{\max(t_{i,m-1}, t) \leq S_i < t_{im}\} \{\mathbf{X}_i \leq \mathbf{x}_i\})}{P(t \leq S_i)}, \quad (17)
\end{aligned}$$

we use (5) and rewrite (17),

$$\begin{aligned}
&\frac{\sum_{j=1}^{n_i+1} I\{t_{i,j-1} \leq t < t_{ij}\}}{1 - P(S_i < t)} \left(\right. \\
&\quad \left. \left(\sum_{m=j}^{n_i} P(\{\max(t_{i,m-1}, t) \leq S_i < t_{im}\} \cap_{k=m}^{n_i} \{a(t_k) + \varepsilon_k \leq x_k\}) \right) \right. \\
&\quad \left. + P(\{\max(t_{i,n_i}, t) \leq S_i < t_{i,n_i+1}\}) \right) \\
&= \frac{1}{1 - F_S(t)} \sum_{j=1}^{n_i+1} I\{t_{i,j-1} \leq t < t_{ij}\} \\
&\quad \sum_{m=j}^{n_i} \int_{s=\max(t_{m-1}, t)}^{t_m} \int \left(\prod_{k=m}^{n_i} \Phi\left(\frac{x_{ik} - a(t_{ik})}{\sigma_\varepsilon}\right) \right) f_{S,C}(s, c) ds dc \\
&\quad + \int_{s=\max(t_{n_i}, t)}^{t_{n_i+1}} \int f_{S,C}(s, c) ds dc. \quad (18)
\end{aligned}$$

We define

$$\Phi\left(\frac{x_{i,n_i+1} - a(t_{i,n_i+1})}{\sigma_\varepsilon}\right) = 1 \quad (19)$$

and rewrite (18) as

$$\begin{aligned}
P(\mathbf{X}_i \leq \mathbf{x}_i \mid a(t) = \eta) &= \frac{1}{1 - F_S(t)} \sum_{j=1}^{n_i+1} I\{t_{i,j-1} \leq t < t_{ij}\} \sum_{m=j}^{n_i+1} \\
&\quad \int_{s=\max(t_{i,m-1}, t)}^{t_{im}} \int \left(\prod_{k=m}^{n_i+1} \Phi\left(\frac{x_{ik} - a(t_{ik})}{\sigma_\varepsilon}\right) \right) f_{S,C}(s, c) ds dc \\
&=: F^{(3)}(\mathbf{x}_i; \theta) \quad (20)
\end{aligned}$$

When $\eta > a_0$ we get

$$P(\mathbf{X} \leq \mathbf{x} \mid a(t) = \eta) = \sum_{j=1}^{n_i+1} \{t_{i,j-1} \leq t < t_{i,j}\} \sum_{m=1}^{n_i+1} \lim_{\epsilon \rightarrow 0} P(\{t_{i,m-1} \leq S_i < t_{im}\} \{\mathbf{X}_i \leq \mathbf{x}_i\} \mid \eta \leq a(t) < \eta + \epsilon). \quad (21)$$

We now use (5) and the fact that we know that $S_i < t$, and rewrite (21)

$$\begin{aligned} & \sum_{j=1}^{n_i+1} \{t_{i,j-1} \leq t < t_{ij}\} \lim_{\epsilon \rightarrow 0} \left[\frac{1}{P(\eta \leq a(t) < \eta + \epsilon)} \left(\sum_{m=1}^{n_i} \right. \right. \\ & \left. \left. P(\{t_{i,m-1} \leq S < \min(t_{im}, t)\} \cap_{k=m}^{n_i+1} \{a(t_{ik}) + \varepsilon_k \leq x_{ik}\} \{\eta \leq a(t) < \eta + \epsilon\}) \right) \right. \\ & \left. + P(\{t_{i,n_i} \leq S < \min(t_{i,n_i+1}, t)\}) \right], \end{aligned} \quad (22)$$

and use definition (19) to include the last term in the summation,

$$\begin{aligned} & \sum_{j=1}^{n_i+1} \{t_{i,j-1} \leq t < t_{ij}\} \sum_{m=1}^{n_i+1} \lim_{\epsilon \rightarrow 0} \\ & \frac{\int_{s=t_{m-1}}^{\min(t_m, t)} \int_{c=\frac{\eta-a_0}{(t-s)^b}}^{\frac{\eta-a_0+\epsilon}{(t-s)^b}} \left(\prod_{k=m}^{n_i+1} \Phi\left(\frac{x_{ik}-a(t_{ik})}{\sigma_\varepsilon}\right) \right) f_{C,S}(c, s) \, ds \, dc}{\int_{s=0}^t \int_{c=\frac{\eta-a_0}{(t-s)^b}}^{\frac{\eta-a_0+\epsilon}{(t-s)^b}} f_{C,S}(c, s) \, ds \, dc} \end{aligned} \quad (23)$$

We denote

$$\tilde{c} = \tilde{c}(s) = \frac{\eta - a_0}{(t - s)^b}, \quad \tilde{a}(t) = a_0 + \tilde{c}(t - s)^b,$$

and

$$\Phi\left(\frac{x_{i,n_i+1} - \tilde{a}(t_{i,n_i+1})}{\sigma_\varepsilon}\right) = 1$$

and rewrite (23),

$$\begin{aligned}
& \sum_{j=1}^{n_i+1} \{t_{i,j-1} \leq t < t_{ij}\} \sum_{m=1}^{n_i+1} \lim_{\epsilon \rightarrow 0} \\
& \frac{\int_{s=t_{i,m-1}}^{\min(t_{im}, t)} \frac{\epsilon}{(t-s)^b} \left(\prod_{k=m}^{n_i+1} \Phi \left(\frac{x_{ik} - \tilde{a}(t_{ik})}{\sigma_\epsilon} \right) \right) f_{C,S}(\tilde{c}, s) ds}{\int_{s=0}^t \frac{\epsilon}{(t-s)^b} f_{C,S}(\tilde{c}, s) ds} \\
& = \frac{\sum_{j=1}^{n_i+1} \{t_{i,j-1} \leq t < t_{ij}\}}{\int_{s=0}^t \frac{1}{(t-s)^b} f_{C,S}(\tilde{c}, s) ds} \\
& \sum_{m=1}^{n_i+1} \int_{s=t_{i,m-1}}^{\min(t_{im}, t)} \frac{1}{(t-s)^b} \left(\prod_{k=m}^{n_i+1} \Phi \left(\frac{x_{ik} - \tilde{a}(t_{ik})}{\sigma_\epsilon} \right) \right) f_{S,C}(s, \tilde{c}) ds \\
& =: F^{(4)}(\mathbf{x}_i; \theta) \tag{24}
\end{aligned}$$

Using (20) and (24) we can write the complete formula as

$$P(\mathbf{X}_i \leq \mathbf{x}_i \mid a(t) = \eta) = I\{a(t) = a_0\}F^{(3)}(\mathbf{x}_i; \theta) + I\{a(t) > a_0\}F^{(4)}(\mathbf{x}_i; \theta). \tag{25}$$

When removing the constraint that $\mathbf{x}_i \geq a_0$ we get

$$\begin{aligned}
& F_{\mathbf{X}_i | a(t)}(\mathbf{x}_i \mid a(t); \theta) = \\
& \left(I\{a(t) = a_0\}F^{(3)}(\mathbf{x}_i; \theta) + I\{a(t) > a_0\}F^{(4)}(\mathbf{x}_i; \theta) \right) I\{\mathbf{x}_i \geq a_0\}, \tag{26}
\end{aligned}$$

and the density function

$$\begin{aligned}
f_{\mathbf{X}_i | a(t)}(\mathbf{x}_i \mid a(t); \theta) & = \sum_{\alpha: |\alpha| + |\alpha'| = n} \partial^\alpha \left(I\{a(t) = a_0\}F^{(3)}(\mathbf{x}_i; \theta) \right. \\
& \left. + I\{a(t) > a_0\}F^{(4)}(\mathbf{x}_i; \theta) \right) I\{\mathbf{x}_i > a_0\}^\alpha \delta_{a_0}(\mathbf{x}_i)^{\alpha'}. \tag{27}
\end{aligned}$$

We now know all the parts of (14), which concludes the derivation of the crack distribution at a fixed time.

3.4 Statistical model

In the previous two sections we derived the distribution of the time until the crack reaches a specific length and the distribution of crack length at a specific time. If the parameters θ are known we can use these distributions to make an inference but since the parameters are estimated from data we have to take the uncertainty in the parameter estimation into account. In this subsection we do this by using a

predictive profile likelihood approach. We also study likelihood based methods for deriving point and interval estimators of the unknown parameters.

The likelihood for the parameters given data of cracks $\underline{\mathbf{x}} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ can be formulated

$$L(\theta) = f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta) = \prod_{i=1}^n f_{\mathbf{x}_i}(\mathbf{x}_i; \theta), \quad (28)$$

where the joint density $f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta)$ is given in (9) ¹.

To get confidence intervals for the parameters in θ we use a profile likelihood approach, cf. Barndorf-Nielsen and Cox [3]. A profile likelihood can be seen as an ordinary likelihood and is generally used when the parameter can be split into two parts $\theta = (\gamma, \xi)$ where ξ is the nuisance parameter part of θ . The profile likelihood and log profile likelihood of γ are defined as

$$\begin{aligned} \tilde{L}(\gamma) &= \sup_{\xi} L(\gamma, \xi), \\ \tilde{l}(\gamma) &= \log \tilde{L}(\gamma), \end{aligned}$$

and the maximum profile likelihood estimate of γ is defined as

$$\hat{\gamma} = \operatorname{argmax}_{\gamma} \tilde{l}(\gamma).$$

From Barndorf-Nielsen and Cox [3] we get

$$2(\tilde{l}(\hat{\gamma}) - \tilde{l}(\gamma)) \xrightarrow{L} \chi_{dim(\gamma)}^2$$

and therefore a $(1 - \alpha)$ confidence region is

$$\tilde{R}_c = \{\gamma : 2(\tilde{l}(\hat{\gamma}) - \tilde{l}(\gamma)) \leq \chi_{dim(\gamma), (1-\alpha)}^2\}. \quad (29)$$

We use (29) to get confidence intervals of the parameters in $\theta = (\psi, \sigma_{\varepsilon}^2, b)$, i.e. if b is the parameter of interest, then the nuisance part is $(\psi, \sigma_{\varepsilon}^2)$

In order to take the parameter uncertainty into account when we make an inference about the time when the crack reaches a fix crack length or the crack length at a fixed time, we use a predictive profile likelihood approach. A predictive profile likelihood for a stochastic variable Y with data $\underline{\mathbf{X}}$ is the likelihood

$$\tilde{L}(y) = \sup_{\theta} L(y, \theta) = \sup_{\theta} f_{Y, \underline{\mathbf{X}}}(y, \underline{\mathbf{x}}; \theta) = \sup_{\theta} \prod_{i=1}^n f_{Y, \mathbf{x}_i}(y, \mathbf{x}_i; \theta). \quad (30)$$

¹Since a product of Dirac measures is ill defined, cf. Schwartz [13], we make a reformulation of the likelihood,

$$f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta) dx = \check{f}_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta) d\check{x} = \check{f}_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta) d(l(a_0) + l)$$

where the Lebesgue measure dx is replaced by a mixed measure $d\check{x}$ containing the point mass measure $l(a_0)$ at the point a_0 and the Lebesgue measure l . The function $\check{f}_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta)$ is similar to $f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \theta)$ but with the Dirac function replaced by the indicator function $I\{x_i = a_0\}$. The same goes for $f_{T|\underline{\mathbf{X}}}(t | \underline{\mathbf{x}}; \theta)$ and $f_{A|\underline{\mathbf{X}}}(a | \underline{\mathbf{x}}; \theta)$.

Normalizing (30) we obtain a density

$$\tilde{f}_Y(y) = \frac{\tilde{L}(y)}{\int \tilde{L}(y) dy},$$

and we view $\tilde{f}_Y(y)$ as an estimate of $f_Y(y)$ and use it for inference about Y .

We now let Y correspond to T given x_i and use (30) to get

$$\tilde{L}(t | \mathbf{x}_i) = \sup_{\theta} f_{\underline{\mathbf{X}}, T | \mathbf{X}_i}(\underline{\mathbf{x}}, t | \mathbf{x}_i; \theta).$$

To avoid the singular distribution $f_{\underline{\mathbf{X}}, T | \mathbf{X}_i}$ we replace it by $f_{\underline{\mathbf{X}}_{-i}, T | \mathbf{X}_i}$. Thus we study instead

$$\begin{aligned} \hat{L}(t | \mathbf{x}_i) &= \sup_{\theta} f_{\underline{\mathbf{X}}_{-i}, T | \mathbf{X}_i}(\underline{\mathbf{x}}_{-i}, t | \mathbf{x}_i; \theta) \\ &= \sup_{\theta} f_{T | \underline{\mathbf{X}}_{-i}, \mathbf{X}_i}(t | \underline{\mathbf{x}}_{-i}, \mathbf{x}_i; \theta) f_{\underline{\mathbf{X}}_{-i} | \mathbf{X}_i}(\underline{\mathbf{x}}_{-i} | \mathbf{x}_i; \theta) \\ &= \sup_{\theta} f_{T | \mathbf{X}_i}(t | \mathbf{x}_i; \theta) f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i}; \theta), \end{aligned} \quad (31)$$

where the equality follows since both T and \mathbf{X}_i are independent of $\underline{\mathbf{X}}_{-i}$, that is cracks on different components grow independent of each other and T is the time until the crack on component i reaches a fixed length a_{max} . The expression $f_{T | \mathbf{X}_i}(t | \mathbf{x}_i; \theta)$ is obtained from the probability model (3) and $f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i}; \theta)$ from (28) via (9).

We now normalize the likelihood (31) to obtain an estimate of the density function

$$\hat{f}_{T | \mathbf{X}_i}(t | \mathbf{x}_i) = \frac{\hat{L}(t | \mathbf{x}_i)}{\int \hat{L}(t | \mathbf{x}_i) dt},$$

and use it for inference of T given \mathbf{x}_i .

If the number of cracks k is large the predictive profile likelihood in (31) is numerically demanding to compute. We can then ignore the uncertainty in the parameter estimation and use $\theta = \hat{\theta}$ for all times t . The likelihood then becomes

$$L(\hat{\theta}, t | \mathbf{x}_i) = f_{T | \mathbf{X}_i}(t | \mathbf{x}_i; \hat{\theta}) f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i}; \hat{\theta}) = f_{T | \mathbf{X}_i}(t | \mathbf{x}_i; \hat{\theta}) K$$

where K is a constant independent of t . Normalization eliminates K and gives the probability model (3) with $\theta = \hat{\theta}$.

If we want to find the distribution of the crack length at a specific time we follow the same steps but use

$$\hat{L}(A | \mathbf{x}_i) = \sup_{\theta} f_{A | \mathbf{X}_i}(a | \mathbf{x}_i; \theta) f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i}; \theta), \quad (32)$$

instead of (31) where $f_{A | \mathbf{X}_i}(a | \mathbf{x}_i; \theta)$ is (14).

Predictive likelihood methods have been examined by e.g. Hinkley [8], Lauritzen [9] and Chib, Jammalamadaka and Tiwari [6]. A review of predictive likelihood methods can be found in Bjornstad [4]. Mathiasen [11] has written more generally about predictive functions. A predictive profile likelihood approach has also been used in Lorén and Lundström [10].

4 Case study

We use the model in the theoretical part to make predictions on crack growth in a low pressure turbine nozzle component. First a small data set is presented and then we discuss a way to estimate the joint distribution of C and S . Finally there is an illustration of the difference of estimating the remaining life with and without taking the uncertainty in the parameter estimation into account.

4.1 Data

The data available is from pri-engines. Pri-engines are engines that are used extensively so they accumulate a large number of flight hours and flight missions. The engines have been observed every 200 flight hours. To illustrate the method, VAC has allowed us to publish crack data from four components. The cracks used to illustrate the model are found in Table 1.

Component	0FH	200FH	400FH	600FH	800FH	1000FH
No1	0	0	0	0	0	-
No2	0	0	0	16	17	-
No3	0	0	25	30	37	-
No4	0	0	20	25	27	33

Table 1: The length in mm of the largest cracks of 4 nozzles observed at 200,400,600,800 and 1000 FH (Flight Hours).

4.2 Estimation of the joint distribution of C and S

In order to use the calculations in Section 3 we need to know the joint distribution of C and S . Figure 1 illustrates the cracks and the crack model (1) fitted to the cracks with a least square method. From each picture we get an observation of S (censored if no crack was detected) and an observation of C if a crack was detected.

Using a similar procedure for all cracks (more than the four in Table 1) indicates that S and C are uncorrelated. We make the assumption that they are independent, that is we assume that the crack growth speed is independent of the time when the crack is initiated. By examining the empirical distribution of S and C we find

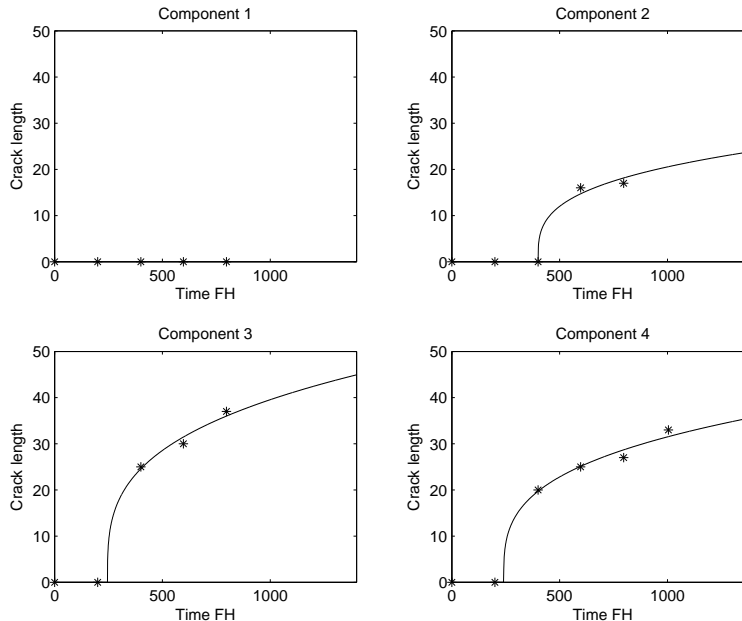


Figure 1: *Model (1) with $a_0 = 0$ fitted to the cracks in Table 1.*

that a log normal distribution is a reasonable approximation. We therefore assume that both S and C are log normally distributed with parameters μ_s, σ_s and μ_c, σ_c , respectively.

4.3 Model illustration

Assume that we have observed the cracks on the three first components and we want to state when the crack on the fourth component reaches length $a_{max} = 30$ mm, that is to find the distribution of T given the observations of component four. We update the distribution T each time we get new observations at 200, 400, 600 and 800 flight hours.

Assume that we know from experience that $\sigma_\varepsilon = 1$ mm but we do not know anything about the other parameters. We use the few observations we have from components one, two and three to estimate the remaining parameters $(\mu_s, \sigma_s, \mu_c, \sigma_c, b)$ using equation (28). First we calculate the distribution of T using (3), ignoring the uncertainty in the parameter estimation. The solid line in Figure 2 illustrates the distribution of T when ignoring parameter estimation uncertainty and $\theta = \hat{\theta}$.

Since the data used to estimate the parameters is limited we can also see how

much the uncertainty in the parameter estimation affects the results. We do this by using the profile likelihood approach (31). The lines with stars in Figure 2 illustrate the distribution of T when we consider the uncertainty of the parameters. The stars indicate where the distribution has been calculated.

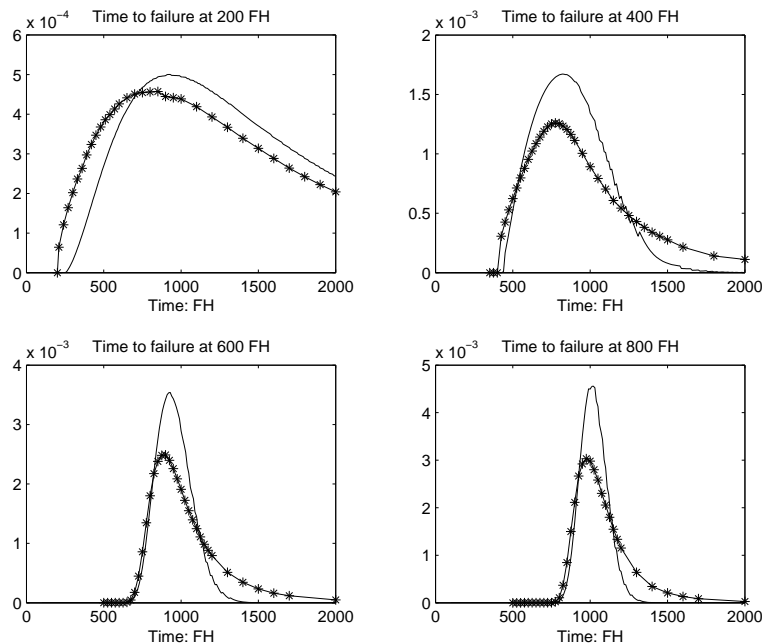


Figure 2: *Distribution of the time when the crack of component four will reach 30mm both considering uncertainty in parameter estimation, line with stars, and not considering uncertainty, solid line. The distribution is updated with the observations at 200 flight hours (upper left), 400 flight hours (upper right), 600 flight hours (lower left) and 800 flight hours (lower right).*

We notice that if we assume that we know the parameters, we seem to be more certain about when the components are going to break than we should be. By plotting both curves we also illustrate the difference. We notice that even if the measurements at 800 flight hours indicate that the component is soon going to break, there is still great uncertainty in the outcome of T . This is due to the fact that each observation has a variance σ_ϵ^2 , and a small change in the values of the crack length results in a large difference in time when the crack length reaches $a_{max} = 30$ mm. This problem originates from the fact that we have decreasing crack growth speed and a measurement variation. Whatever way is chosen to analyze the crack we will get a

similar problem. The same principal works to our advantage if we wish to predict the crack length at a fixed time.

This can be illustrated by using the theory in section 3.3 where we find the crack length distribution at a fix time. If we do this analysis for all possible times we get a three-dimensional picture of how the crack will grow over time given the information in our observations, illustrated in Figure 3. We can update the picture as we get

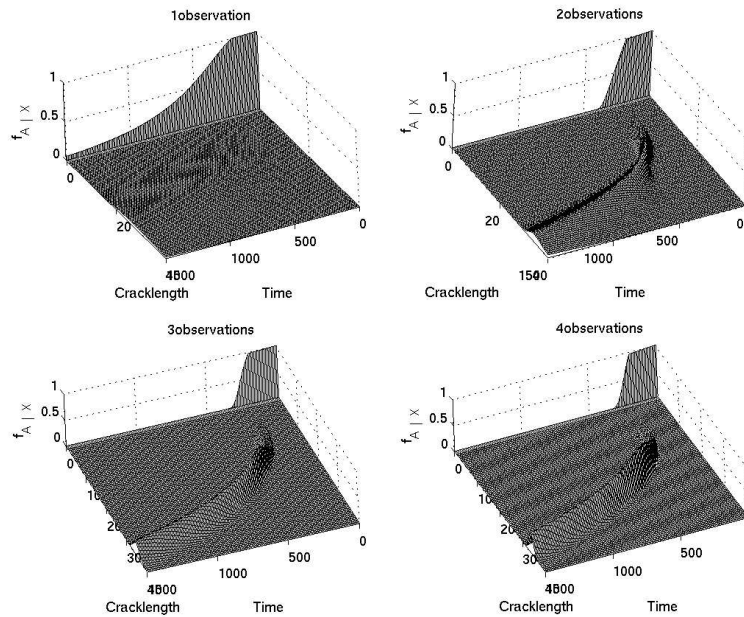


Figure 3: *Three-dimensional illustration of how the distribution of the crack length at a fixed time of component four changes over time. The distribution is updated each time a new observation is made: first observation at 200 flight hours (upper left), second observation at 400 flight hours (upper right), third observation at 600 flight hours (lower left) and forth observation at 800 flight hours (lower right). The values at crack length 0 correspond to the probability point mass that there are no cracks.*

more information and we notice that we get more and more certain as to how the crack will behave in the future as well as how the crack has behaved in the past. This type of picture is good if we want to evaluate the risk of an unplanned engine maintenance occurrence due to the fact that the crack length exceeds a_{max} before it is expected to. Note that if we integrate the function in Figure 3 with regard to crack length we will get marginal function of value 1 for all times. The high values in the upper part of the pictures are the point mass that indicates that the crack

length is below size a_0 . In the case of one observation, the upper left picture, we can observe how the probability of a crack length of length a_0 is decreasing as the time increases. At 0 FH this probability is one. The observation that we have no crack at 200FH means there is a very slim chance that there is a crack at this time. Hence the probability is almost one. In the other pictures in Figure 3 we get a feeling for that even if we are fairly certain of the crack length at a fixed time the distribution of the time when the crack reaches a specific crack length will have a large variance.

If we combine the probability model in Section 3.3 with the statistical profile likelihood based approach in Section 3.4, we get a comparison of how the uncertainty in the parameters affect the distribution of crack length at a fixed time. We plot this for times 500FH and 1000FH as illustrated in Figure 4 where we also update the illustration when we get more observations.

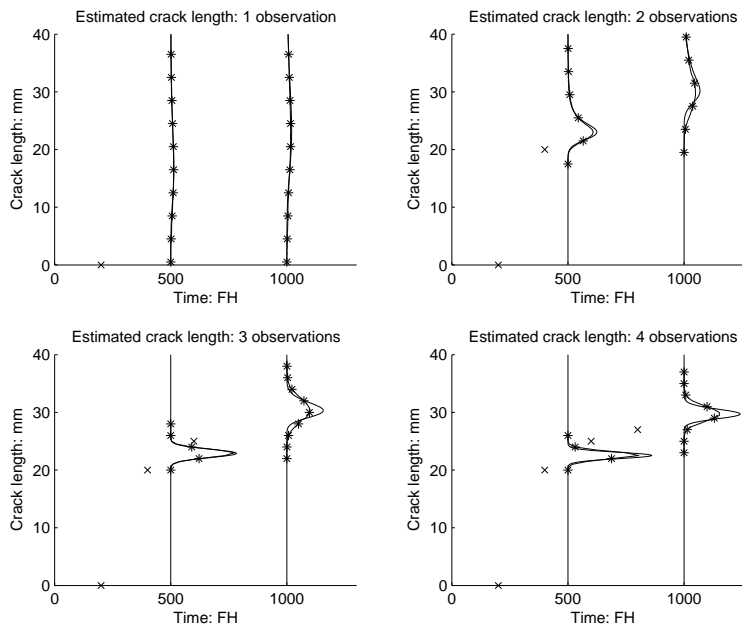


Figure 4: *Distribution of the crack length at times 500 and 1000 flight hours of component four both considering uncertainty in parameter estimation, line with stars, and not considering uncertainty, solid line. The distributions are updated with the observations at 200 flight hours (upper left), 400 flight hours (upper right), 600 flight hours (lower left) and 800 flight hours (lower right). The observations are marked in pictures with "x".*

5 Summary

We have used real data and made an empirical model describing the crack growth on an air craft engine nozzle component. With the help of the model we have presented formulas to answer the question when a component is going to break. A component is considered broken when the crack reaches a specific length a_{max} . We have found the distribution of the time when a crack will reach length a_{max} , given earlier observations of the crack. From the model we also get the distribution of the crack length at a fixed but arbitrary time. This distribution is updated when more observations are available. The distributions can be calculated both if we know the parameters in the model and if we have to estimate them. If we have to estimate the parameters we use a profile likelihood approach to take the uncertainty of the parameter estimates into account when we calculate the distributions. Whatever way we choose to estimate the distribution of time when the crack reaches length a_{max} , it will get a relatively large variation. This is due to the decreasing crack growth speed and the measurement variation.

The methods illustrated can be used to predict a variety of other failures that are not related to crack growth, e.g. corosions failures, even if the choice of empirical model is different. The main strength of an empirical approach compared to a mechanistic approach is that hard environments can be modeled without a complete understanding of the physics involved. The weakness is that the models may not be moved, can not be extrapolated outside the observation space and may perform incorrectly if the load characteristic changes.

Acknowledgements

The author wants to thank Dragi Anevski and Jacques de Maré at Chalmers and Fredrik Plyhm at VAC for there supervision. This work was financed by NFFP (Nationella Flygtekniska Forsknings Programmet) and VAC.

References

- [1] N. Andréasson. Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems. Licentiate thesis, Chalmers University of Technology, 2004.
- [2] Tom M. Apostol. *Mathematical Analysis*. Addison Wesley Longman, 1974.
- [3] O. E. Barndorff-Nielsen and D. R. Cox. *Inference and asymptotics*, volume 52 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1994.

- [4] Jan F. Bjørnstad. Predictive likelihood: a review. *Statist. Sci.*, 5(2):242–265, 1990. With comments and a rejoinder by the author.
- [5] V.V. Bolotin. Reliability against fatigue fractures in the presence of set of cracks. *Engineering Fracture Mechanics*, 53:753–759, 1996.
- [6] Siddhartha Chib, S. Rao Jammalamadaka, and Ram C. Tiwari. A new definition of the predictive likelihood. *Statist. Probab. Lett.*, 5(2):113–118, 1987.
- [7] D.B. W. Cutshall Garcia, J.W Cardinal, C.B Fohn, O.H Burnside, and J. Pendleton. Application of fracture mechanics and damage tolerance techniques to a nose landing gear drag brace fitting. *Presented at the International Conference: Residual Fatigue Life and Life Extension of In-Service Structures.*, 11th International Spring Meeting(June 2006), Paris.
- [8] David Hinkley. Predictive likelihood. *Ann. Statist.*, 7(4):718–728, 1979.
- [9] Steffen L. Lauritzen. Sufficiency, prediction and extreme models. *Scand. J. Statist.*, 1(3):128–134, 1974.
- [10] S. Lorén and M. Lundström. Modelling curved s-n curves. *Fatigue & Fracture of Engineering Materials and Structures*, 28(5):437–443, 2005.
- [11] P. E. Mathiasen. Prediction functions. *Scand. J. Statist.*, 6(1):1–21, 1979.
- [12] Yudi Pawitan. *In all likelihood*. Clarendon press, Oxford, 2001.
- [13] Laurent Schwartz. Sur l'impossibilité de la multiplication des distributions. *C. R. Acad. Sci. Paris*, 239:847–848, 1954.
- [14] D.C. Virkler, B.M Hillberry, and P.K Goel. The statistical nature of fatigue crack propagation. *Journal of Engeneering Materials and Technology*, 101:148–153, 1979.
- [15] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139:469–489, 2002.
- [16] J.N. Yang. A simple second order approximation for stochastic crack growth analysis. *Engineering Fracture Mechanics*, 53:677–686, 1996.

Paper D

Discrete Approximations of Life Distributions in Optimal Replacement

Johan Svensson
Chalmers University of Technology
Department of Mathematical Statistics
and
Göteborgs University
Department of Mathematical Statistics
SE-41296 Göteborg, Sweden

Abstract

Discretization of continuous random variables is used in a stochastic optimization problem. We present a measure connected to an optimization model that suggests what parts to replace in an aircraft engine. The optimization model requires a discrete random variable with restrictions on the points of support. Four methods to make discretizations are discussed and adapted to the constraints of the model. The methods are compared and the choice of the number of points of support is discussed. Finally the consequence of using a so called narrow scenario tree is commented upon.

Keywords: optimal maintenance; discretization; points of support

1 Introduction

Aircraft engines can be more economically maintained and resources saved if the maintenance process is optimal. The optimality is here defined in economical terms, and the main factor for economical savings determines which components should be replaced on each service occasion. Several optimization models have been developed to deal with this problem. Epstein and Wilamowsky [5] and Dickman et al. [3] have developed models for components with predetermined deterministic lives. Andréasson [2] has developed a model for details with deterministic lives but also details with stochastic lives. Another approach to the problem is to construct a general maintenance policy that is not always optimal but at least satisfactory. A survey of replacement and maintenance polices can be found in Wang [14].

We will use the model presented in Andréasson [2], formulate it as a two-stage model, and define an error measure. Different discrete approximations will be used in order to incorporate components with stochastic lives into the model. With the help of the error measure we will compare different approximations. Work on different ways of making discretizations of continuous distribution functions for implementation in models has been done by Høyland and Wallace [7], Keefer [9], Miller and Rice [10] and Smith [12], [13]. Other work on discretization and tree reduction in multi-stage problems has been done by i.e Dupačová et al. [4], Frauendorfer and Schürle [6] and Pflug [11]. We will make discretizations with restrictions on the points of support of the density function.

This paper consists of three parts. In the first part we present the optimization model and the error measure. The second part introduces different discretizations methods, and the last part compares the different methods with respect to the error measure. The objective is to conclude how many points of support are necessary and what method of discretization to use. Furthermore, we discuss the consequence of making a narrow scenario tree.

2 Optimization model

The aim of the optimization model is to minimize the expected cost of maintenance for an aircraft engine. In order to describe the maintenance we classify the components of the engines into two categories; deterministic details and stochastic details.

Definition 1: A *deterministic detail* is a component that has a predetermined limited time in service which must not be exceeded.

Components that are vital for safety are deterministic. If a deterministic component fails there is a risk that the engine will stop functioning. The predetermined time limit is set so low that there is practically no risk the component will fail before this time.

Definition 2: A *stochastic detail* is a component that is not deterministic.

Stochastic components are allowed to operate in the aircraft engine until they fail. If a stochastic component fails during a flight the engine will still work but with reduced performance. In this paper only one stochastic detail is used but it is possible to extend the model to several stochastic details. The life U of a new stochastic component is modeled with a distribution \tilde{G} and the remaining life of a functioning

stochastic component with an age of u_0 is modeled with a distribution G where

$$\begin{aligned} G(u) &= P(U \leq u + u_0 \mid U > u_0) = \frac{P(u_0 < U \leq u + u_0)}{1 - P(U \leq u_0)} \\ &= \frac{\check{G}(u + u_0) - \check{G}(u_0)}{1 - \check{G}(u_0)}. \end{aligned}$$

We assume that $\check{G}'(u) > 0$ if $u > 0$ and U to be a non-negative random variable.

The engine is repaired for two reasons, either the stochastic detail has failed or a deterministic detail has reached its predetermined time limit. The repair of the engine is associated with a cost for bringing the engine to service. At service, there is an opportunity to replace other components and consequently extend the time to the next repair. Every time the engine is at the repair bay, an optimization is performed. The input data is the state of the components in the engine and the output data is a suggestion of what parts to replace. The optimization model is described in detail in Andréasson [2].

In order to solve the optimization problem, for an engine consisting of many parts, simplifications must be made. The main simplification in this paper is that in the sequence of life distributions for the stochastic detail only the first life distribution is modeled in a stochastic way, the remaining life distributions are replaced by the expected value of the life distribution. This simplification makes the model easier to formulate and solve, but there are some drawbacks, described in Section 4.2.

Another simplification inherent in the optimization model is that the time, during which the maintenance cost is minimized, is divided into T discrete time points instead of a continuous approach. The optimization model is formulated as a stochastic two-step model. More information about stochastic optimization models can be found in Kall and Wallace [8].

2.1 First stage model

In order to decide which components to replace we introduce the first step binary variables x_0^i for deterministic details $i = 1, \dots, N$ and \tilde{s}_0 for the stochastic detail. The variables $x_0^1, \dots, x_0^N, \tilde{s}_0$ take the values 0 or 1 with a 1 indicating replacement of the detail. The lower index indicates the time point and zero indicates that this is a decision we have to make now, hence a first step variable. We introduce the replacement strategy vector $\mathbf{x}_0 = (x_0^1, \dots, x_0^N, \tilde{s}_0)$ and we want to find the solution

$$\hat{\mathbf{x}}_0 = \operatorname{argmin}_{\mathbf{x}_0 \in \{0,1\}^{N+1}} F(\mathbf{x}_0),$$

with

$$F(\mathbf{x}_0) = \int_0^\infty f(\mathbf{x}_0, u) dG(u) = E_G[f(\mathbf{x}_0, U)], \quad (1)$$

where f is defined in section 2.2.

2.2 Second stage model

In the second stage model we take into consideration possible future problems. Overall we want to minimize the expected cost of maintaining the engine during a fix time period containing T equidistant time points. We denote the time points nodes. One approximation of the optimization model is that components are only allowed to break and be replaced at nodes. If T is large we get good resolution at the cost of long calculation time, and if T is smaller the resolution gets worse but the calculation time shortens.

The lives of the deterministic components that are currently in the engine correspond to the node indices $\tilde{\tau}_1, \dots, \tilde{\tau}_N$.

The life of the stochastic component currently in the engine U is transformed to a node by $\tilde{\tau}_s = \tilde{\tau}_s(u)$, which is a function from U to an index set $1, \dots, T$ describing which node corresponds to every u . For later convenience let us construct $\tilde{\tau}_s(u)$ as a step function in such a way that it is right continuous, increasing in u , and does not jump at any nodes. Furthermore if U takes the exact time corresponding to node i , we let $\tilde{\tau}_s(u) = i$. This implies that f (defined in (2) below) for a fixed \mathbf{x}_0 is a right continuous step function in u and does not jump at any node. The reason for this is that f only changes values when $\tilde{\tau}_s(u)$ is changing values. The function f also decreases in u (formulated as a lemma in section 2.4). The function $\tilde{\tau}_s$ can be written as

$$\tilde{\tau}_s(u) = i \text{ if } u \in [w_{i-1}u_{i-1} + (1 - w_{i-1})u_i, w_i u_i + (1 - w_i)u_{i+1}),$$

where u_i is the time corresponding to node i and $w_i \in (0, 1)$. This definition means that we treat components that are going to break some time after u_i as broken at time u_i and we therefore replace them at time u_i . This is an approximation since at time u_i we never know if the components will break in the near future. The motivation for this approach is that we are interested in the first stage variables \mathbf{x}_0 and not the replacement schemes (the second stage variables) if we get the outcome u . Remember here that the nodes do not correspond to any actual maintenance times. In reality a stochastic component is replaced as soon as it is observed as broken, whether this times corresponds to a node or not.

The lives of new (replacing) deterministic and stochastic details are described with node indices τ_1, \dots, τ_N and τ_s with τ_s indicating the node closest to the expected value of the life distribution \check{G} .

We introduce the second stage binary variables as $(x_t^1, \dots, x_t^N), t = 1, \dots, T$ for the deterministic components, $\tilde{s}_t, t = 1, \dots, T$ for the stochastic component currently in the engine, and $s_t, t = 1, \dots, T$ for the replacing stochastic components. Furthermore, we let z_t be binary variables indicating whether service is performed on the engine at time $t = 1, \dots, T$. Thus the vectors $(x_t^1, \dots, x_t^N, \tilde{s}_t, s_t, z_t)$ describe if replacement and service are performed at times $t = 1, \dots, T$ with 1 indicating replacement and service, respectively.

Costs associated with the maintenance are c_i, c_s , and d , where c_i is the cost of

replacing component $i = 1, \dots, N$, c_s the cost of replacing the stochastic component and d a fixed cost for the service process. Thus $(c_1, \dots, c_N, c_s, d)$ is the vector of costs associated with the replacements and service made at each node.

In order to simplify the notation we introduce the set $\mathcal{N} = \{1, \dots, N\}$. The second stage function f describes the cost if we make replacement \mathbf{x}_0 , when the stochastic life is u and is defined as

$$f(\mathbf{x}_0, u) = \min_{\mathbf{x} \in \psi} \sum_{t=0}^T \left(\sum_{i \in \mathcal{N}} c_i x_t^i + c_s (s_t + \tilde{s}_t) + dz_t \right), \quad (2)$$

where ψ is the set of points $\mathbf{x} = (x_t^1, \dots, x_t^N, s_t, \tilde{s}_t, z_t; t = 1, \dots, T)$ such that

$$x_0^i + \sum_{t=1}^{\tilde{\tau}_i} x_t^i \geq 1, \quad i \in \mathcal{N}, \quad (3)$$

$$\sum_{t=\ell}^{\tau_i + \ell - 1} x_t^i \geq 1, \quad \ell = 0, \dots, T - \tau_i, \quad i \in \mathcal{N}, \quad (4)$$

$$x_t^i \leq z_t, \quad t = 0, \dots, T, \quad i \in \mathcal{N}, \quad (5)$$

If $\tilde{\tau}_s(u) \geq T$ constraints (6)-(13) shall be removed

$$\sum_{t=0}^T \tilde{s}_t = 1, \quad (6)$$

$$\sum_{t=0}^{\tilde{\tau}_s} \tilde{s}_t = 1, \quad (7)$$

$$\sum_{t=\ell}^{\tau_s + \ell - 1} s_t + \sum_{t=\ell}^{\tilde{\tau}_s(u)} \tilde{s}_t \geq 1, \quad \ell = 0, \dots, \min(\tilde{\tau}_s(u), T - \tau_s), \quad (8)$$

$$\sum_{t=\ell}^{\tau_s + \ell - 1} s_t \geq 1, \quad \ell = \tilde{\tau}_s(u) + 1, \dots, T - \tau_s, \quad (9)$$

$$s_0 = 0, \quad (10)$$

$$s_t \leq \sum_{k=0}^{t-1} \tilde{s}_k, \quad t = 1, \dots, \tilde{\tau}_s(u), \quad (11)$$

$$\tilde{s}_t \leq z_t, \quad t = 0, \dots, \tilde{\tau}_s(u), \quad (12)$$

$$s_t \leq z_t, \quad t = 0, \dots, T, \quad (13)$$

$$z_0 = 1, \quad (14)$$

$$x_t^i, s_t, \tilde{s}_t, z_t \in \{0, 1\}, \quad t = 0, \dots, T, \quad i \in \mathcal{N}. \quad (15)$$

The constraints (3) force the installed deterministic components to be replaced before their lives $\tilde{\tau}_i$ are consumed. The components that replace the current deterministic components are not allowed to be in the engine more than τ_i nodes. This is regulated by constraints (4). Constraints (5) force the indicator variables z_t to be one if we replace any deterministic component at node t .

Constraints (6) to (13) are constraints regulating replacement of the stochastic component and should be removed if $\tilde{\tau}_s(u) \geq T$, because then replacement of the stochastic component is not necessary. Constraint (6) tells us that the installed stochastic component can only be replaced once and (7) tells us that the replacement must occur before node $\tilde{\tau}_s$. The component that replaces the current stochastic component is not allowed to be in the engine more than τ_s nodes. This is regulated by constraints (8) and (9).

Constraints (10) and (11) mean that no replacing stochastic component is allowed to be installed before the current stochastic component is removed.

Constraints (12) and (13) force the z_t variables to be one if we replace the stochastic component at node t . The engine is at the repair bay at time zero, hence constraint (14) sets $z_0 = 1$. In constraints (15) we have the binary restrictions.

Note that in this model it is only the time to first failure $\tilde{\tau}_s$ of the stochastic component that is modeled with distribution G . The remaining times are all assumed (modeled) to be deterministic and equal to the node τ_s closest to expectation under G . This is sometimes called a narrow scenario tree, see Altenstedt [1]. In the model there is no requirement that the components function at node T .

2.3 Maximum discretization

In the above formulation $\tilde{\tau}_s$ corresponds to a discretization of G . The structure of the second stage model requires discretization to work. In section 2.4 we will measure how good a discretization is. In order to do that we introduce two kinds of discretizations. The first discretization allows probability mass on all nodes and the second allows probability mass only at a subset of the nodes.

Let $\kappa_T = \{u_1, \dots, u_T\}$ be the maximal set of nodes where probability mass is allowed, let $n < T$ be a positive integer, and denote $\kappa_n = \{k_1, \dots, k_n\} \subset \kappa_T$.

Definition 3: The discretization G_T is a discretization of G that allows probability mass in *all* T nodes in κ_T . We call this discretization the *maximum discretization*.

Definition 4: The discretization G_n is a discretization of G that allows probability mass in at most n nodes in κ_T . We call this discretization the *n-node discretization*.

A maximum discretization G_T has the corresponding probability mass function

$$g_T(u) = \begin{cases} p_{1_T}, & \text{if } u = u_1, \\ \vdots & \vdots \\ p_{T_T}, & \text{if } u = u_T, \end{cases} \quad (16)$$

and a n-node discretization G_n the probability mass function

$$g_n(u) = \begin{cases} p_1, & \text{if } u = k_1, \\ \vdots & \vdots \\ p_n, & \text{if } u = k_n. \end{cases} \quad (17)$$

We discuss different discretization approaches in section 3.

Nodes with non-zero probability mass are also called points of support for a distribution. Points of support for a distribution $H(u)$ are all points u such that, if $a < u < b$, then $H(b) - H(a) > 0$, for any a, b .

2.4 Measure of error

In order to reduce the time needed to complete the optimization, we want a discretization with as few points of support as possible but we still want a good replacement strategy. Using the maximum discretization (16) or the n-node discretization (17) yields different replacement strategies. The maximum discretization gives the replacement strategy

$$\hat{\mathbf{x}}_0^T = \underset{\mathbf{x}_0 \in \{0,1\}^{N+1}}{\operatorname{argmin}} F_T(\mathbf{x}_0), \quad (18)$$

where $F_T(\mathbf{x}_0)$ is defined in (1) with G replaced by G_T , so that

$$F_T(\mathbf{x}_0) = \sum_{i=1}^T f(\mathbf{x}_0, u_i) p_{i_T}.$$

The n-node discretization (17) gives the replacement strategy

$$\hat{\mathbf{x}}_0^n = \underset{\mathbf{x}_0 \in \{0,1\}^{N+1}}{\operatorname{argmin}} F_n(\mathbf{x}_0), \quad (19)$$

with F_n defined in (1) with G replaced by G_n , so that

$$F_n(\mathbf{x}_0) = \sum_{i=1}^n f(\mathbf{x}_0, u_i) p_i.$$

As G_T is a distribution using the maximum number of nodes in the model we compare the quality of G_n to G_T by introducing the error measure for the expected cost between two discretizations G_T and G_n as

$$e(G_n, G_T) = F_T(\hat{\mathbf{x}}_0^n) - F_T(\hat{\mathbf{x}}_0^T). \quad (20)$$

Note that $e(G_n, G_T) \geq 0$. In order to make a discretization that has a small error we use the following result.

Theorem 1: The error measure can be bounded with the following inequalities,

$$e(G_n, G_T) \leq 2 \sup_{\mathbf{x}_0} |F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0)| \leq C \sup_u |G_n(u) - G_T(u)|, \quad (21)$$

where C is a bounded constant.

The first inequality is proven in Pflug [11] but the short proof is stated here for the convenience of the reader.

Proof:(First inequality) Set $\varepsilon = \sup_{\mathbf{x}_0} |F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0)|$. Let $M = \{\mathbf{x}_0 : F_T(\mathbf{x}_0) \leq F_T(\hat{\mathbf{x}}_0^T) + 2\varepsilon\}$. Suppose that $\hat{\mathbf{x}}_0^n \notin M$, then

$$F_T(\hat{\mathbf{x}}_0^T) + 2\varepsilon < F_T(\hat{\mathbf{x}}_0^n) \leq F_n(\hat{\mathbf{x}}_0^n) + \varepsilon \leq F_n(\hat{\mathbf{x}}_0^T) + \varepsilon \leq F_T(\hat{\mathbf{x}}_0^T) + 2\varepsilon.$$

This contradiction establishes $\hat{\mathbf{x}}_0^n \in M$, i.e.

$$e(G_n, G_T) = F_T(\hat{\mathbf{x}}_0^n) - F_T(\hat{\mathbf{x}}_0^T) \leq 2\varepsilon = 2 \sup_{\mathbf{x}_0} |F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0)|. \square$$

In the second inequality in (21) we use the fact that the second stage function f in (1) is right continuous and does not jump at any nodes. We also use the following lemma.

Lemma 1: The second stage function f decreases in u for any fix \mathbf{x}_0 , i.e

$$f(\mathbf{x}_0, u) \geq f(\mathbf{x}_0, u + \varepsilon), \quad \forall \varepsilon > 0$$

Proof: We will check that the point $\mathbf{x} = (x_t^1, \dots, x_t^N, s_t, \tilde{s}_t, z_t; t = 1, \dots, T)$, which solves the minimization problem (2) when $u = u_0$ still satisfies the constraints when $u = u_0 + \varepsilon$. This means that $f(\mathbf{x}_0, u)$ decreases in u since we can guarantee the same cost at $u = u_0 + \varepsilon$ as when $u = u_0$.

The relevant constraints to consider are (7),(8),(9) and (11). In the case when $\tilde{\tau}_s(u_0) = \tilde{\tau}_s(u_0 + \varepsilon)$, all constraints are the same. In the case when $\tilde{\tau}_s(u_0) < T$

and $\tilde{\tau}_s(u_0 + \varepsilon) \geq T$, \mathbf{x} belongs to ψ when $u = u_0 + \varepsilon$ since removing constraints means less restriction on the second stage variables. We now consider the case when $\tilde{\tau}_s(u_0) < \tilde{\tau}_s(u_0 + \varepsilon) < T$.

We know from constraints (6) and (7) that \mathbf{x} satisfies $\tilde{s}_t = 0$, $t \geq \tilde{\tau}_s(u_0) + 1$. This observation makes (7) true when $u = u_0 + \varepsilon$.

We now consider constraints (8) and (9) when $u = u_0 + \varepsilon$ but $\tilde{s}_t = 0$, $t \geq \tilde{\tau}_s(u_0) + 1$. The first $\tilde{\tau}_s(u_0) + 1$ constraints in (8) are identical to the constraints in (8) when $u = u_0$. The remaining $\tilde{\tau}_s(u_0 + \varepsilon) - \tilde{\tau}_s(u_0)$ constraints in (8) are identical to the first $\tilde{\tau}_s(u_0 + \varepsilon) - \tilde{\tau}_s(u_0)$ constraints in (9) when $u = u_0$. All the constraints in (9) when $u = u_0 + \varepsilon$ can be found in (9) when $u = u_0$.

The first $\tilde{\tau}_s(u_0)$ constraints in (11) are identical when $u = u_0$ and $u = u_0 + \varepsilon$. When $u = u_0 + \varepsilon$ and we use $\tilde{s}_t = 0$, $t \geq \tilde{\tau}_s(u_0) + 1$ the remaining $\tilde{\tau}_s(u_0 + \varepsilon) - \tilde{\tau}_s(u_0)$ constraints state

$$s_t \leq \sum_{i=1}^{\tilde{\tau}_s(u_0)} \tilde{s}_i, \quad t = \tilde{\tau}_s(u_0) + 1, \dots, \tilde{\tau}_s(u_0 + \varepsilon),$$

which is not really a restriction since $\sum_{i=1}^{\tilde{\tau}_s(u_0)} \tilde{s}_i = 1$ according to constraint (7), and all variables are binary according to (15).

We have now checked that solution \mathbf{x} does not violate any constraints when $u = u_0 + \varepsilon$ since the constraints do not mean any restrictions or are identical to the ones occurring when $u = u_0$. \square

A consequence of Lemma 1 is that the longer the life of the stochastic detail is the cheaper it is to maintain the engine.

Proof:(Second inequality): Denote $Y = f(\mathbf{x}_0, U) = f(U)$ and note that Y is a non-negative stochastic variable. We rewrite

$$P(Y > y) = 1 - P(f(U) \leq y) = 1 - P(U \geq f^{-1}(y)) = P(U < f^{-1}(y)),$$

where

$$f^{-1}(y) = \begin{cases} \infty, & \text{if } y < f_{min}, \\ \inf\{z : f(z) \leq y\}, & \text{if } f_{min} \leq y \leq f_{max}, \\ 0, & \text{if } y > f_{max}, \end{cases}$$

and f_{min} and f_{max} depend on \mathbf{x}_0 and are the minimum and maximum costs of maintaining the engine. Note that $P(U < f^{-1}(y)) = G(f^{-1}(y))$ if f does not change values at node points, i.e. if $f^{-1}(y) \notin \kappa_T$. Note that this follows from the assumption on $\tilde{\tau}_s$.

Equation (1) can be rewritten

$$\begin{aligned} E_G[Y] &= \int_0^\infty P(Y > y) dy = \int_0^\infty P(U < f^{-1}(y)) dy \\ &= \int_0^\infty G(f^{-1}(y)) dy = f_{min} + \int_{f_{min}}^{f_{max}} G(f^{-1}(y)) dy. \end{aligned}$$

Now

$$\begin{aligned} & 2 \sup_{\mathbf{x}_0} | F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0) | \\ &= 2 \sup_{\mathbf{x}_0} | E_{G_T}[Y] - E_{G_n}[Y] | \\ &= 2 \sup_{\mathbf{x}_0} \left| \int_{f_{min}}^{f_{max}} G_T(f^{-1}(y)) - G_n(f^{-1}(y)) dy \right| \\ &\leq C \sup_u | G_T(u) - G_n(u) |, \end{aligned}$$

with $C = 2 \sup_{\mathbf{x}_0} (f_{max} - f_{min})$. Then

$$\begin{aligned} f_{max} \leq \hat{f}_{max} = \min \left\{ \sum_{i \in \mathcal{N}} \left(\left\lceil \frac{T}{\tau_i} \right\rceil + 1 \right) (c_i + d) + \left(\left\lceil \frac{T}{\tau_s} \right\rceil + 1 \right) (c_s + d), \right. \\ \left. T d + \sum_{i \in \mathcal{N}} \left(\left\lceil \frac{T}{\tau_i} \right\rceil + 1 \right) c_i + \left(\left\lceil \frac{T}{\tau_s} \right\rceil + 1 \right) c_s \right\}, \end{aligned}$$

and

$$f_{min} \geq \hat{f}_{min} = \left\lceil \frac{T}{\max_i(\tau_i)} \right\rceil d + \sum_{i \in \mathcal{N}} \left\lceil \frac{T}{\tau_i} \right\rceil c_i,$$

where $[x]$ is the integer part of x . Thus C is bounded. \square

3 Different discretization approaches

When modeling a discretization of $G(u)$, $u \in [0, \infty)$ with $n \leq T$ points of support, the following questions arise:

1. How many points of support should we use?
2. Which points of support $\kappa_n \subset \kappa_T$ should we choose?
3. How should we place the probability mass?

Answering questions 2 and 3 simultaneously may lead to optimization problems, that are as difficult to solve as the original optimization problem. In section 4 we try to answer question 1 by simulation.

We will describe four different approaches and describe how existing methods can be adapted to our situation. The methods are first presented without any restrictions and then with our restrictions.

3.1 Method minimizing the Sup-distance

Theorem 1 bounds the error measure (20) by a constant times the sup-distance between G_n and G_T . Here we discuss a discretization that minimizes the sup-distance between G_n and G . The reason is that when T tends to infinity, G_T tends to G uniformly in order to be a sequence of discrete approximations of G . The triangle and inverse triangle inequalities gives

$$\begin{aligned} \sup_u |G_n(u) - G(u)| &\leq \sup_u |G_n(u) - G_T(u)| + \sup_u |G_T(u) - G(u)| \\ &= \sup_u |G_n(u) - G_T(u)| + \varepsilon, \end{aligned}$$

and

$$\begin{aligned} \sup_u |G_n(u) - G(u)| &\geq \sup_u |G_n(u) - G_T(u)| - \sup_u |G_T(u) - G(u)| \\ &= \sup_u |G_n(u) - G_T(u)| - \varepsilon, \end{aligned}$$

which gives

$$\sup_u |G_n(u) - G_T(u)| - \varepsilon \leq \sup_u |G_n(u) - G(u)| \leq \sup_u |G_n(u) - G_T(u)| + \varepsilon,$$

where $\varepsilon = \sup_u |G_T(u) - G(u)|$ is small when T is large.

Assume that we do not have a restriction that the points of support have to be in the set κ_T . Then we can minimize $\sup_u |G(u) - G_n(u)|$ where G_n is a discretization with n points of support with masses p_i at u_i , $i = 1 \dots, n$ by choosing point of support i as

$$u_i = G^{-1}\left(\frac{2i-1}{2n}\right) \quad \text{and} \quad p_i = \frac{1}{n}. \quad (22)$$

This can be realized by looking at Figure 1, where a,b,c,d,e and f indicate the greatest difference between G and G_n in intervals 1,2,3 and 4. The overall sup-distance is equal to the maximum distance of a,b,c,d,e or f and is minimized if all distances a,b,c,d,e and f are equal. This gives a sup-distance of $(2n)^{-1}$.

With our restrictions on the points of support we get a greater sup-distance. Assume we fix the n points of support in κ_T and form discretization (17). Let

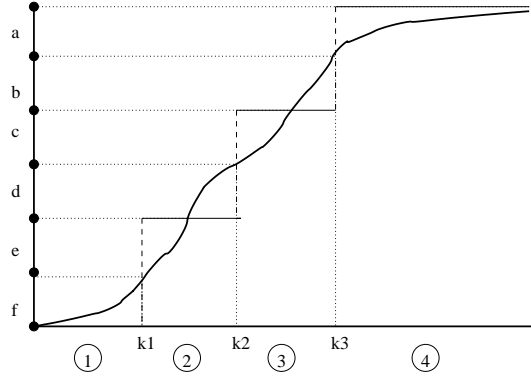


Figure 1: A fictive distribution G and G_3 where a, b, c, d and f are the maximum sup-distance between the functions in intervals 1, 2, 3 and 4.

$\mathbf{p} = (p_1, \dots, p_n)$ be the probability vector and $\hat{\mathbf{p}}$ the solution to

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \{ \sup_u |G(u) - G_n(u)| \}. \quad (23)$$

In most cases $\hat{\mathbf{p}}$ is not unique. We then choose the solution that minimizes the sup-distance in every interval $[(0, k_1), (k_1, k_2), \dots, (k_n, \infty)]$. The overall sup-distance is the maximum of the sup-distance in those intervals so the solution also minimizes the overall sup-distance (23). The solution is as follows

$$\begin{aligned} p_1 &= \frac{G(k_1) + G(k_2)}{2}, \\ p_i &= \frac{G(k_i) + G(k_{i+1})}{2} - \frac{G(k_{i-1}) + G(k_i)}{2} = \frac{G(k_{i+1}) - G(k_{i-1})}{2}, \\ p_n &= 1 - \frac{G(k_n) + G(k_{n-1})}{2}, \end{aligned}$$

where $i = 1, \dots, n-1$. In the intervals $(0, k_1)$ and (k_n, ∞) the sup-distance will be $G(k_1)$ and $1 - G(k_n)$ respectively regardless of $\hat{\mathbf{p}}$. In interval (k_i, k_{i+1}) the minimum sup-distance is $(G(k_{i+1}) - G(k_i))/2$ and it is achieved since

$$\begin{aligned} G_n(k_i) &= \sum_{j=1}^i p_j = \frac{G(k_1) + G(k_2)}{2} + \sum_{j=2}^i \frac{G(k_{j+1}) - G(k_{j-1})}{2} \\ &= \frac{G(k_i) + G(k_{i+1})}{2}, \end{aligned}$$

and

$$\sup_{u \in (k_i, k_{i+1})} |G_n(k_i) - G(u)| = \frac{G(k_{i+1}) - G(k_i)}{2}.$$

If we have the liberty to choose both p_i and which n points of support $k_i \in \kappa_T$ we want to use we get the following problem. Let $(p_i, k_i), i = 1, \dots, n$ be the discretization with p_i the mass in node k_i . Let $\hat{\mathbf{p}}, \hat{\kappa}_n$ be the solution to

$$(\hat{\mathbf{p}}, \hat{\kappa}_n) = \operatorname{argmin}_{\mathbf{p}, k_i \in \kappa_T} \{ \sup_u |G(u) - G_n(u)| \}. \quad (24)$$

As above the solution is not unique. Solving this problem is the same as solving problem (23) for all $\binom{T}{n}$ choices of points of support. (Many combinations are unlikely to be best and can be omitted.) In general a good set of points of support are points that lie dense where G increases much.

Instead of the above optimization we suggest to choose points of support near the points in (22). This can be achieved by choosing the n points of support as follows.

1. Choose the first point of support as

$$k_1 = \operatorname{argmin}_{u \in \kappa_T} |G(u) - \frac{1}{2n}|.$$

2. Choose the i 'th point of support as

$$k_i = \operatorname{argmin}_{u \in \kappa_T \cap u > k_{i-1}} |G(u) - \frac{2(1 - G(k_{i-1}))}{2(n - i + 1) + 1} - G(k_{i-1})|. \quad (25)$$

Here we try to spread $G(k_1), \dots, G(k_n)$ uniformly in the interval $[0, 1]$ but since k_i has to be from set κ_T this is not always possible. Equation (25) considers the remaining part of the interval $[0, 1]$, namely $[G(k_{i-1}), 1]$ and spreads the points in it. Here we start from the beginning of the interval $[0, 1]$, but it would be equally good to start from the end. Note that if there is no restriction of possible points, this choice coincides with (22). Now probabilities can be chosen by (23).

3.2 Bracket means method

This method consists of dividing G into n intervals $\{[t_0, t_1], \dots, [t_{n-1}, t_n]\}$ where $t_0 = 0, t_n = \infty$. We now make a discretization by putting the mass

$$p_i = \int_{t_{i-1}}^{t_i} dG(t),$$

at the point

$$u_i = \frac{\int_{t_{i-1}}^{t_i} t dG(t)}{\int_{t_{i-1}}^{t_i} dG(t)}.$$

The intervals can be chosen in many ways. One common choice is that all points of support have equal probability mass. A similar method is the bracket median method, in which one uses the median in each subinterval instead of the mean, see Smith [13].

In the case where the points of support have to be in the set κ_T , it is not always possible to find subintervals so the mean in each subinterval corresponds with a node in κ_T . A simple example that shows this is this: One interval $[0, \infty)$ and mean in distribution does not match node in κ_T .

We present two ways to find approximations that almost satisfy the bracket mean condition. For the first approach we decide which points of support we should use. In the second approach we use a probability vector that helps to choose points. Both methods are expressed as optimization problems that are rewritten in a standard form in appendix A.

Approach 1: Let μ_1 measure the distance between point sets. Fix the n preferred nodes $\kappa_n \subset \kappa_T$. Let $\tilde{\kappa}_n = \{\tilde{u}_1, \dots, \tilde{u}_n\}$ be a set of points of support, that is not a subset of κ_T . The solution $\hat{\mathbf{p}} = \{p_1, \dots, p_n\}$ is obtained for the optimization problem

$$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \phi}{\operatorname{argmin}} \mu_1(\kappa_n, \tilde{\kappa}_n) \quad (26)$$

where ϕ is the set of points \mathbf{p} such that

$$\begin{aligned} p_i &= \int_{t_{i-1}}^{t_i} dG(t), \quad i = 1, \dots, n, \\ \tilde{u}_i &= \frac{1}{p_i} \int_{t_{i-1}}^{t_i} t dG(t), \quad i = 1, \dots, n, \\ t_i &> t_{i-1}, \quad i = 1, \dots, n, \\ t_0 &= 0, \\ t_n &= \infty. \end{aligned}$$

Solving (26) gives the discretization of the distribution as in (17). A possible choice of μ_1 is $\mu_1 = \sum_{i=1}^n w_i (k_i - \tilde{u}_i)^2$ where w_i is a weight.

Approach 2: Let μ_{21} measuring the distance between probability vectors and μ_{22} measuring the distance between point sets. Let $\mathbf{q} = \{q_1, \dots, q_n\}$ be a probability vector with desirable probabilities and let $\tilde{\kappa}_n = \{\tilde{u}_1, \dots, \tilde{u}_n\}$ be a set of points of support, that is not a subset of κ_T . Then a discretization $\mathbf{p} = \{p_1, \dots, p_n\}$ with points of support $\kappa_n = \{k_1, \dots, k_n\}$ is obtained as the solution $(\hat{\mathbf{p}}, \hat{\kappa}_n)$ to the optimization problem

$$(\hat{\mathbf{p}}, \hat{\kappa}_n) = \underset{\mathbf{p} \in \phi, \kappa_n \subset \kappa_T}{\operatorname{argmin}} \mu_{21}(\mathbf{q}, \mathbf{p}) + \mu_{22}(\kappa_n, \tilde{\kappa}_n), \quad (27)$$

where ϕ is the same set as in the first approach. The solution of (27) gives the discretization of the distribution as in (17). Measures μ_{21} and μ_{22} need to be chosen so that they do not completely dominate each other. If μ_{21} dominates, we get the brackets with the probabilities we desire but the mean in each bracket may be far from a node. If μ_{22} dominates, the mean in each bracket will be near a node but the probabilities will be far from those desired.

3.3 Method minimizing Wasserstein distance

Pflug [11] suggests the following discretization

$$\begin{aligned} p_1 &= \int_0^{\frac{k_1+k_2}{2}} dG(t), \\ p_i &= \int_{\frac{k_{i-1}+k_i}{2}}^{\frac{k_i+k_{i+1}}{2}} dG(t), \quad i = 2, \dots, n-1, \\ p_n &= \int_{\frac{k_{n-1}+k_n}{2}}^{\infty} dG(t), \end{aligned}$$

where $\{k_1, \dots, k_n\} \in \kappa_n$ are the points of support of the given G_n , derived as the discrete distribution minimizing the Wasserstein distance, see Pflug [11]. In Pflug [11] an error measure similar to equation (20) is used, for which he presents a bound depending on the Lipschitz continuity of f with respect to u . His result is not directly applicable to our problem since f is not Lipschitz continuous but we will nevertheless use this method to decide probabilities for comparison purposes.

3.4 Moment preserving method

If we have no constraints on the points of support and the first $2n-1$ moments of the distribution G are finite, then it is possible to create a discrete approximation with n points of support that correctly matches $2n-1$ moments. Let

$$M_j = \int_{-\infty}^{\infty} u^j dG(u),$$

be the j 'th moment. The discretization can be obtained by finding u_i and p_i that satisfy

$$\sum_{i=1}^n p_i u_i^j = M_j \quad j = 0, \dots, 2n-1. \quad (28)$$

For a solution see Miller and Rice[10] and Smith [12]. It can be shown that if all M_j are finite and from a probability distribution that spans $[a, b]$ then all u_i will be real and lie in the interval $[a, b]$ and all $p_i \geq 0$.

In our problem, we must choose points from the set κ_T . In order to compare the methods we will use the same points as for the method that minimizes the sup-distance.

Another approach is to calculate points of support using (28) and then choose the points of support in κ_T that are closest to the points obtained. For some distributions this method resulted in points very far out in the tail.

We have n fixed points of support $u_i, i = 1, \dots, n$ and we want to find probabilities $p_i, i = 1, \dots, n$ so that the discretization matches as many moments as possible. Following a simple argument of degrees of freedom we see that the maximum number of moments we can approximate is $n - 1$, (if M_0 is not counted as a moment). Thus we look for a solution p_1, \dots, p_n to

$$\sum_{i=1}^n p_i u_i^j = M_j \quad j = 0, \dots, n - 1. \quad (29)$$

There is no guarantee that there exists a solution with $p_i \geq 0, i = 1, \dots, n$. If some $p_i < 0$ we suggest removing one constraint. The least important constraint is usually the highest order moment constraint

$$\sum_{i=1}^n p_i u_i^{n-1} = M_{n-1}.$$

Removing this we obtain the equations (29) with $j = 0, \dots, n - 2$. If the solution to these satisfies $p_i \geq 0, i = 1, \dots, n$ we are done. If not, we should remove one more moment constraint and keep removing constraints until a solution is found resulting in $p_i \geq 0, i = 1 \dots n$.

After removing moment constraints the solution to the problem is not necessarily unique. (There are more variables than equations). In order to choose one solution we can use a function z that represents some other desired properties in the discretization and solve

$$\hat{\mathbf{p}} = \operatorname{argmax}_{\mathbf{p} \in \varphi} z(\mathbf{p}),$$

where φ is the set of points $\mathbf{p} = (p_1, \dots, p_n)$ such that

$$\begin{aligned} \sum_{i=1}^n p_i u_i^j &= M_j, \quad j = 0, \dots, n - 1 - m, \\ p_i &\geq 0, \quad i = 1, \dots, n, \end{aligned}$$

where m is the number of removed moment constraints. The function z can, e.g., be formulated to promote p_i of the same sizes if that is a desired property.

4 Test results

In this section we will present numerical results illustrating the different discretization methods in section 3. We will use a Weibull distribution to describe the life of the stochastic component. We use the following parameterization,

$$\check{G}(u) = 1 - e^{-(\frac{1}{\theta}u)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0) \quad (30)$$

where θ is the characteristic life and α is the shape parameter. The maximum number of time nodes, T , was set to 30. The distance between the time nodes was set to one and θ was set to 9. Tests were made with the α parameter being both 1 and 2.

In the following we will make an attempt to establish a rule of thumb for how many points of support are needed. We will also comment on the consequence of using a narrow scenario tree.

4.1 Error measure

We study how the error measure depends on different discretizations methods and different number of points of support. We model the engine with two components, one stochastic and the other deterministic. When determining components to replace, there are four alternatives:

1. Replace the deterministic component.
2. Replace the stochastic component.
3. Replace both components.
4. Do not replace any components.

The optimal replacement alternative was calculated in optimization problem (19) with $n = 1, \dots, 10$ points of support.

The points of support were chosen by equation (25). The probabilities were chosen in four ways, by the method that minimizes the sup-distance, the method that minimizes the Wasserstein distances, the method that preserves the moments, and the bracket method approach 1.

The best discretization possible was a discretization with 30 points of support, one in every node, for which the optimal replacement alternative was calculated according to (18). Finally the difference between the two discretizations, using error measure (20), was calculated. The calculation was performed in AMPL (a modeling language for mathematical programming).

The parameters and remaining lives of the components used are shown in Table 1.

The expected life of the stochastic component τ_s is connected to the shape parameter α in (30). The parameter τ_s was set to 9 when $\alpha = 1$ and 8 when $\alpha = 2$. The

Parameter	Alternatives
$\tilde{\tau}_1$	4, 6
τ_1	6, 10
c_1	60, 100, 130
c_s	70, 100, 150
d	70, 100, 150
α	1, 2

Table 1: Parameters changed in the optimization model (2) and Weibull distribution (30).

age of the stochastic component was changed from new to an age of 9 in four steps, (new, 3, 6, 9). The error measure was calculated for all levels of the parameters, ages of the stochastic component and discretization methods. In Figure 2 the result of the test is presented.

The method that preserves the moments seems worse and the methods that minimize the Wasserstein and Sup-distance seem better. It seems that using two points of support is worse than just using the expected value of the distribution. At three points of support or more the error measure seems rather constant, compared with using one or two points of support, if we disregard the moment method with $\alpha = 1$. Some work has been done on three points of support discrete distributions, cf. Keefer [9].

Further tests showed that the error measure decreases as α increases. With constant θ , the variance of the Weibull distribution (30) decreases with the increasing α parameter. If the variance is great it is harder to describe the distribution with just a few points of support.

4.2 Narrow tree approximation

The greatest approximation is that only the first life of the stochastic component is modeled with several points of support. After the stochastic component is replaced it is modeled with the expected value of the life distribution, which is the same as using one point of support. The reason why this approximation was invented in the first place was that it was natural to think that what happens in the near future has a greater impact on the decision we have to make today, than what is going to happen a long time from now. The optimization model works in a different way. Basically it tries to find a scheme with replacement times that, for the entire service period, minimize the expected cost of maintaining the engine. The approximations sometimes lead to irrational solutions, such as when a new stochastic detail is replaced.

A simple example that illustrates this is the following one. Assume we have one deterministic and one stochastic detail. The engine is at the repair bay because the deterministic detail needs replacement. For the sake of simplicity assume that the

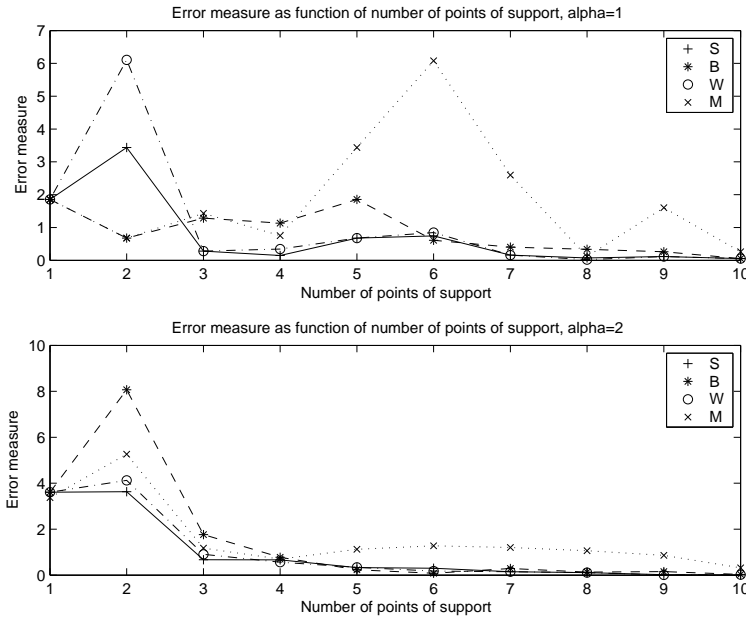


Figure 2: Y-axis is the mean error measure with parameters as in Table 1. X-axis is the number of points of support, S = method that minimizes the sup-distance, B = Bracket method, W = method that minimize Wasserstein distance, M = Moment preserving method.

stochastic component is new, that is $\check{G} = G$. We use the model to answer the question if we should replace the stochastic detail. Assume the price for making repairs is $z_t = 20$ and that replacing a component costs $c_1 = c_s = 10$. We model the stochastic component with 2 points of support. It fails in node 3 with probability $p_A = 0.5$ and in node 5 with probability $p_B = 0.5$. The expected value of \check{G} corresponds to a distance of 4 nodes. Consequently the stochastic component that replaces the first stochastic component must be replaced at least at every four nodes. Assume that the deterministic component has a life of 4 nodes. The total time we need to maintain the engine is $T = 8$ nodes. Note that there is no restriction that the engine needs to function at node 8. In figure 3 we see three replacement schemes.

In node zero we see the first stage variables, that is the decision we make now. Given that we have to replace the deterministic component we can either keep the stochastic component (schemes A and B) or replace it (scheme C). For each scheme a cost q is calculated. The model will suggest a replacement of the stochastic component if the cost $q_C < p_A q_A + p_B q_B$. With our prices this is $80 < 90$. Consequently

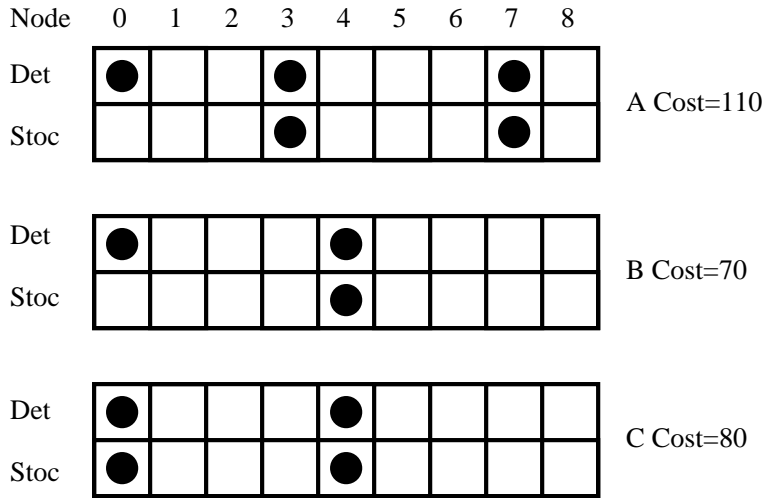


Figure 3: *Three replacement schemes in an optimization problem that contains one deterministic and one stochastic component. A dot indicates replacement. The prices of replacing a component is 10 and the cost of bringing the engine to service is 20*

we replace the stochastic component even if it is assumed to be new.

In the case where $\alpha = 1$ the Weibull distribution (30) becomes an exponential distribution. The exponential distribution has a constant failure or hazard rate, which means that the risk of failure is constant over time. Thus a stochastic component with an exponential distribution never needs replacement. Table 2 contains results on the percent of replacements from the test run with the exponential distribution. It seems that the model very often replaces the component even if replacement is not necessary. The problem is not solved by increasing the number of points of support. When 30 points of support is used the stochastic component is still replaced in 18% of the times. When only one point of support is used there is no replacement of the stochastic component because the replacing component is modeled in the same way. This model problem can probably be corrected if more than the first stochastic life distribution is modeled with several points of support.

5 Summary and further work

An optimization model for replacement of parts in an aircraft engine containing stochastic and deterministic components has been studied. We made a discretization of the distribution of the life of the stochastic component. The structure of the optimization model demands that the points of support of the discretization coincide with

PoS	S	B	W	M
1	0	0	0	0
2	0.44	0.11	0.56	0.11
3	0.10	0.31	0.10	0
4	0.15	0.20	0.19	0.06
5	0.11	0	0.11	0.44
6	0.31	0.22	0.31	0.55
7	0.24	0.27	0.24	0.41
8	0.12	0.09	0.13	0.24
9	0.24	0.26	0.24	0.37
10	0.13	0.13	0.13	0.26

Table 2: The percent of times the optimization model replaced the stochastic component even if it was not aging. PoS=Points of Support, S = method that minimizes Sup-distance, B = Bracket method, W = method that minimizes Wasserstein distance, M = Moment preserving method.

the nodes in the model. Four different discretization methods were presented, the bracket mean method, the moment preserving method, a Wasserstein distance minimizing method, and a method that minimizes the Sup-distance. The discretization methods were adapted to the constraints in the optimization model.

An error measure closely connected to the optimization model was defined and a maximum limitation of the error was derived. With the help of the error measure and test runs the different methods were compared. According to the test, the methods that minimize the Wasserstein and Sup-distance were better. The moment preserving method performed worse. The minimum number of points of support suggested is three.

When using the narrow scenario tree many details were replaced even if they did not need replacement. This model problem can probably be corrected if more than the first stochastic life distribution is modeled with several points of support.

Possible future work is to model more lives with several points of support. Furthermore we need to study a model with more stochastic details and how to make discretizations if the lives of the components are correlated.

Acknowledgements

The author wants to thank Dragi Anevski at Göteborg University and Jacques de Maré at Chalmers for supervision. This work was financed by NFFP.

Appendix

A: Bracket method. The optimization problems in section 3.2 is here rewritten in a less explicit form. Let

$$H(\mathbf{t}) = \begin{pmatrix} H_1(t_1, t_2) \\ \vdots \\ H_i(t_{i-1}, t_i) \\ \vdots \\ H_n(t_{n-1}, t_n) \end{pmatrix} = \begin{pmatrix} \int_{t_0}^{t_1} dG(u) \\ \vdots \\ \int_{t_{i-1}}^{t_i} dG(u) \\ \vdots \\ \int_{t_{n-1}}^{t_n} dG(u) \end{pmatrix}, \quad (31)$$

$$B(\mathbf{t}) = \begin{pmatrix} B_1(t_1, t_2) \\ \vdots \\ B_i(t_{i-1}, t_i) \\ \vdots \\ B_n(t_{n-1}, t_n) \end{pmatrix} = \begin{pmatrix} \frac{\int_{t_0}^{t_1} t dG(u)}{H_1(t_1, t_2)} \\ \vdots \\ \frac{\int_{t_{i-1}}^{t_i} t dG(u)}{H_i(t_{i-1}, t_i)} \\ \vdots \\ \frac{\int_{t_{n-1}}^{t_n} t dG(u)}{H_n(t_{n-1}, t_n)} \end{pmatrix}, \quad (32)$$

where $\mathbf{t} = (t_1, \dots, t_n)$. Problem (26), (fixed κ_n) can now be formulated

$$\bar{\mathbf{t}} = \operatorname{argmin}_{\mathbf{t} \in \Gamma} \mu_1(\kappa_n, B(\mathbf{t})), \quad (33)$$

and the problem (27) and be formulated

$$(\bar{\mathbf{t}}, \bar{\kappa}_n) = \operatorname{argmin}_{\kappa_n \subset \kappa_T, \mathbf{t} \in \Gamma} \mu_{21}(\mathbf{p}, H(\mathbf{t})) + \mu_{22}(\kappa_n, B(\mathbf{t})), \quad (34)$$

where Γ is the set of points \mathbf{t} such that

$$\begin{aligned} t_i &\geq t_{i-1} + \epsilon, \quad i = 1, \dots, n, \\ t_0 &= 0, \\ t_n &= G^{-1}(p_x), \end{aligned}$$

where ϵ is a positive small number and p_x is a probability near 1. If μ_{21} and μ_{22} are continuous the objective function will be continuous. The constraints form a compact set in \mathbf{t} . Problem (33) is solved by standard methods. Problem (34) can be solved for every fix choice of $\kappa_n \subset \kappa_T$. We can make $\binom{T}{n}$ different choices of κ_n and need to check which one minimizes the objective function in (34). For large T this is computationally demanding.

References

- [1] F. Altenstedt. *Aspects on asset liability management via stochastic programming*. PhD thesis, Chalmers University of Technology, 2003.
- [2] N. Andréasson. Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems. Licentiate thesis, Chalmers University of Technology, 2004.
- [3] Jitka Dupačová, Giorgio Consigli, and Stein W. Wallace. Scenarios for multi-stage stochastic programs. *Ann. Oper. Res.*, 100:25–53 (2001), 2000. Research in stochastic programming (Vancouver, BC, 1998).
- [4] Y. Epstein, S. Wilamowsky. Opportunistic replacement in a deterministic environment. *Comput. and Opns. Res.*, 12:311–322, 1985.
- [5] B. Dickman et al. Modeling deterministic opportunistic replacement as an integer programming problem. *Amer. J. Math. and Mgmt.*, 11, 1991.
- [6] Karl Frauendorfer and Michael Schürle. Term structure models in multistage stochastic programming: estimation and approximation. *Ann. Oper. Res.*, 100:189–209 (2001), 2000. Research in stochastic programming (Vancouver, BC, 1998).
- [7] W. Høyland, K. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.
- [8] Peter Kall and Stein W. Wallace. *Stochastic programming*. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons Ltd., Chichester, 1994.
- [9] D.L. Keefer. Certain equivalents for three-point discrete-distribution approximations. *Management Science*, 40:760–773, 1994.
- [10] T.R. Miller, A.C. Rice. Discrete approximations to probability distributions. *Management Science*, 29:352–362, 1983.
- [11] G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program.*, 89(2, Ser. B):251–271, 2001. Mathematical programming and finance.
- [12] J.E. Smith. *Moment method for decision analysis*. PhD thesis, Stanford University, 1990.
- [13] J.E. Smith. Moment methods for decision analysis. *Management Science*, 39:340–358, 1993.
- [14] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139:469–489, 2002.