CHALMERS | UNIVERSITY OF GOTHENBURG

PREPRINT 2008:45

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

TORGNY ALMGREN
NICLAS ANDRÉASSON
DRAGI ANEVSKI
MICHAEL PATRIKSSON
ANN-BRITH STRÖMBERG
JOHAN SVENSSON

Department of Mathematical Sciences
Division of Mathematics
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Göteborg Sweden 2008

Preprint 2008:45

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

Department of Mathematical Sciences
Division of Mathematics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg, Sweden
Göteborg, December 2008

Preprint 2008:45
ISSN 1652-9715

Matematiska vetenskaper
Göteborg 2008

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 study that indicates that a non-stationary renewal process with Weibull distributed lives is a good model for the recurring maintenance occasions. Using 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.

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., water and 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. [9].)

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

*Volvo Aero Corporation, SE-461 81, Trollhättan, Sweden.
†Tegnérsgatan 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.

we consider the safety-critical parts as having deterministic life limits. These limits are measured in "cycles" (to be defined below) 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.

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.

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

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 [11]). 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 corresponding 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, 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.

In order to remove a specific part from a module it is most often necessary to remove others 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. Here, 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.

1.1 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.

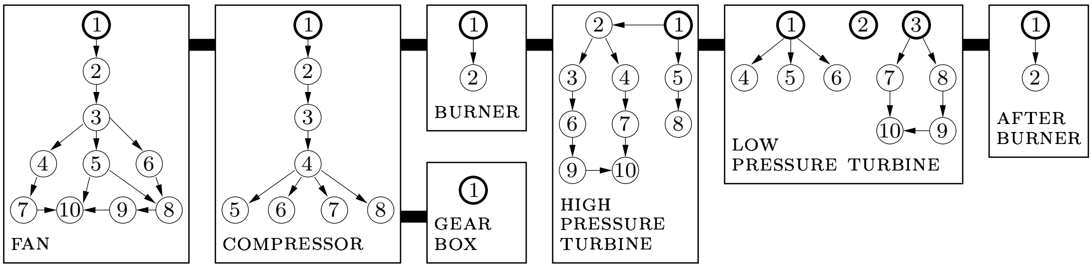

Figure 1: *A graphical representation of the deterministic parts of the RM12 engine. Each box represents a module, each node part, and each arc a possible path to reach the part towards which the arc is pointing.*

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 ([9]) 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 model is designed to consider the cost for production interruptions while minimizing the cost of maintenance. In practice the means that the optimization 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.

1.2 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). Contracts however often describe how the difference in engine status, between the start and end of the 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. 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. The earlier a customer signs a new contract the better the maintenance activities can be planned; a reasonable policy is to give the customer some type of discount if a new contract is signed before the current contract runs out.

1.3 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, 9].) Under an *age replacement policy* (e.g., [4]) a component is replaced at failure or at a specified age, whichever occurs first. Age replacement policies can also be governed by condition monitoring devices, or be based on fixed intervals (e.g., the block replacement policy in [4]). Their main drawback is that practically new items may be replaced at planned replacement times. (See [5] and below for modifications dealing with this issue.) 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. Age and block replacement policies are examples of *scheduled maintenance policies*. They 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 in [14].

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.

In [12] the assumption is that 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; when it fails it destroys the whole system but it has no associated cost for loss of performance. The authors conclude that optimal policies are likely to be extremely difficult to compute and also 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. In [13] the maintenance of the compressor of an aircraft engine is considered. 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.)

1.4 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. Related literature mainly 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. 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 the necessary ones 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 the 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 $t = 0, 1, \ldots, T \geq 2$. At time step $t = 0$ all of the parts of the system are new and at $t = T$ the system will be discarded. We define $\mathcal{N} = \{1, \ldots, N\}$. 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. The objective is to minimize the cost of having a working system between the time steps 0 and T.

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

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

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

The variables z_t and x_{it} are not defined for $t \in \{0, T\}$, since it is not beneficial to replace any part at these times. Each part has a life limit $T_i \leq T-1$; it must be replaced at least once every T_i time steps, which is forced by the constraints

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

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

$$x_{it} \leq z_t, \qquad i \in \mathcal{N}, \quad t = 1, \ldots, T-1. \tag{2}$$

The model presented in [10] includes the constraints $\sum_{i \in \mathcal{N}} x_{it} \leq N z_t$, $t = 1, \ldots, T-1$, instead of (2), which are stronger in the sense that the set defined by (2) is included in that defined by these constraints.

A complete model of the minimization of the total cost for having a working system between the time steps 0 and T is called the *replacement problem* and is given by the problem to

$$\text{minimize} \quad \sum_{t=1}^{T-1} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + d z_t \right), \quad \text{subject to} \quad (x, z) \in S, \tag{3}$$

where

$$S = \left\{ (x, z) \in \mathbb{B}^{N(T-1)} \times \mathbb{B}^{T-1} \;\middle|\; (x, z) \text{ fulfils (1) and (2)} \right\}. \tag{4}$$

Example 1 (numerical illustration) We 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 is solved for $d = 0$, 10, and 1000 (where $d = 10$ represents the most reasonable value in a real maintenance situation). For $d = 0$, the optimal total number of replacement occasions is 11 and there is no advantage with replacing a component before its life limit is reached. Increasing d from 0 to 10 decreases the optimal total number of replacement occasions 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. For $d = 1000$ it is very important to utilize the opportunity to replace several components at a time. The optimal total number of replacement occasions is four (the least feasible number of replacement occasions for this instance).

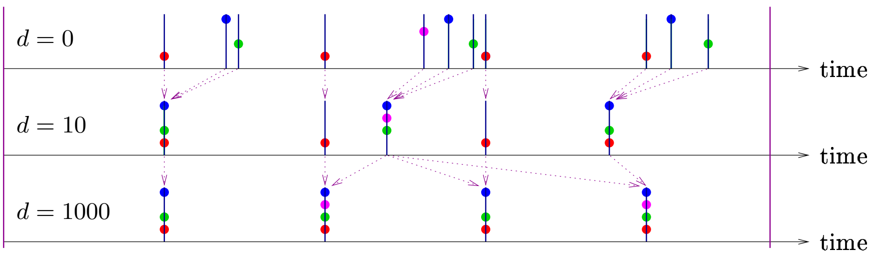

Figure 2: *Optimal maintenance schedules for $d = 0$, 10, and 1000. When d increases from 0 to 10 the replacement occasions 1–3, 5–7, and 9–11, are grouped into one occasion for each of the three groups. When d is increased from 10 to 1000, the last four maintenance occasions are rearranged into three occasions; the reduction from five to four occasions results in several more component replacements.*

Figure 2 shows the maintenance occasions for the three cases. The horizontal axis represents the 60 time steps 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. □

3 The replacement polytope

We study the structure of the set S, defined in (4), of feasible solutions to (3). The convex hull of S, denoted $\operatorname{conv} S$, is called the *replacement polytope*. By studying the facial structure of S and thereby describe its convex hull by a finite set of linear inequalities, it is 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 most of the inequalities in (4) define facets of the same. However, by an example we show that these basic inequalities do not completely define $\operatorname{conv} S$. We then derive a new class of facets by using Chvátal–Gomory rounding.

3.1 Polyhedral combinatorics

We review the results on polyhedral combinatorics necessary for deriving our results on the facial structure of the replacement polytope. A comprehensive survey of polyhedral combinatorics is given in [15].

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 $x^1, \ldots, x^m \in \mathbb{R}^n$ if there exist scalars $\lambda_1, \ldots, \lambda_m$ with $\lambda_1 + \cdots + \lambda_m = 1$ such that $x = \lambda_1 x^1 + \cdots + \lambda_m x^m$. The *affine hull* of X, denoted by $\operatorname{aff} X$, 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 $x \in \operatorname{aff}(X \setminus \{x\})$. 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 $P = \{ x \in \mathbb{R}^n \mid Ax \leq b \}$, 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 dimension of P is given by $\dim(P) + \operatorname{rank}(A^=, b^=) = n$ ([15, p. 87]). If $\dim P = n$ we say that P is full-dimensional.

If V is a finite set in $\mathbb{R}^n$ and $X = \operatorname{conv} V$, then each extreme point of X lies in V ([20, p. 81]). Moreover, every polytope equals the convex hull of its extreme points ([7, p. 206]), and an obvious relation is then that, if $V \subseteq \mathbb{R}^n$ then $\dim V = \dim(\operatorname{conv} V)$. Another useful result is that a set is a polytope if and only if it is a bounded polyhedron ([20, p. 114]).

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. If $A \in \mathbb{R}^{m \times n}$ is a totally unimodular matrix and $b \in \mathbb{R}^m$ is integral, then the polyhedron defined by $Ax \leq b$ is integral ([7, p. 221]). We will utilize the following characterization of total unimodularity.

Proposition 2 (characterization of the TU property, [15, 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 $J \subseteq \{1, \ldots, n\}$ there exists a partition $\{J_1, J_2\}$ of J such that the inequality $\left|\sum_{s \in J_1} a_{rs} - \sum_{s \in J_2} a_{rs}\right| \leq 1$ holds for $r = 1, \ldots, m$.* □

Let the polyhedron P be given as above. 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. [15, p. 89]) that if F is a facet of P, then there exists some affine inequality defining F.

A full-dimensional polyhedron P has a unique (to within scalar multiplication) minimal representation by a finite set of affine inequalities ([15, p. 91]), and 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 $P = \{ x \in \mathbb{R}^n \mid a^i x \leq b_i,\ i = 1, \ldots, k \}$, where $a^i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$.

From the previous it follows that if $V \subseteq \mathbb{R}^n$ is a finite set, then the polytope $\operatorname{conv} V$ is a polyhedron. Hence, if $\operatorname{conv} V$ is full-dimensional, it then follows that the union of all facet-defining inequalities of $\operatorname{conv} V$ defines an affine 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, is useful when proving that a certain valid inequality defines a facet.

Proposition 3 ([15, 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 $\alpha \in \mathbb{R}$.* □

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

3.2 The dimension and basic facets of $\operatorname{conv} S$ and a new class of facets

In this section we derive the dimension of the replacement polytope $\operatorname{conv} S$ and investigate the inequalities used to define S in (4). Under weak and natural assumptions we show that the replacement polytope is full-dimensional. Further, we show that all inequalities that are necessary to define the replacement polytope are facets of the same.

Lemma 4 *The polyhedron defined by* (1) *and*

$$-x_{it} \geq -1, \qquad i \in \mathcal{N}, \quad t = 1, \ldots, T-1, \tag{5}$$

is integral. □

Proof We derive the result by showing that the constraint matrix of (1), (5) is TU using the characterization in Proposition 2. The inequalities (1), (5) separate over $i \in \mathcal{N}$; therefore it suffices to show that the constraint matrix of the inequality system

$$\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1, \qquad \ell = 1, \ldots, T - T_i, \tag{6a}$$

$$-x_{it} \geq -1, \qquad t = 1, \ldots, T-1, \tag{6b}$$

is TU for each $i \in \mathcal{N}$. Let $A^i \in \mathbb{B}^{(T-T_i)\times(T-1)}$ be the constraint matrix defined by the left hand sides of the inequalities (6a), that is, for each $r \in \{1, \ldots, T - T_i\}$, let $a^i_{rs} = 1$ for $s \in \{r, \ldots, T_i + r - 1\}$ and $a^i_{rs} = 0$ for $s \in \{1, \ldots, T-1\} \setminus \{r, \ldots, T_i + r - 1\}$. The essential property of the matrix A^i is that the ones appear consecutively in each row, that is, if $a^i_{r\ell} = a^i_{rk} = 1$ and $1 \leq \ell \leq k \leq T-1$, then $a^i_{rs} = 1$ for all $s \in \{\ell, \ldots, k\}$; this property is closed under column deletions. Let $B \in \mathbb{B}^{(T-1)\times(T-1)}$ be the constraint matrix defined by the left hand sides of the inequalities (6b). Then $B = -I^{T-1}$ (minus the identity matrix); 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 property (ii) of Proposition 2 is satisfied for $J = \{1, \ldots, T-1\}$. Let $J_1 = \{j \in J \mid j \text{ odd}\}$ and $J_2 = J \setminus J_1$. For each $\ell \in \{1, \ldots, T - T_i\}$ it holds that

$$\sum_{s \in J_1} a^i_{\ell s} - \sum_{s \in J_2} a^i_{\ell s} = \begin{cases} 1, & \text{if } T_i \text{ is odd and } \ell \text{ is odd}, \\ -1, & \text{if } T_i \text{ is odd and } \ell \text{ is even}, \\ 0, & \text{if } T_i \text{ is even}, \end{cases}$$

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

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

It follows that the property (ii) stated in Proposition 2 holds. Hence, the constraint matrix $((A^i)^{\mathrm{T}}, B^{\mathrm{T}})^{\mathrm{T}}$ of (6) is TU. Since the right-hand sides of (6) are all integral it follows from [7, p. 221] that the corresponding polyhedron is integral. □

Proposition 5 (dimension of the replacement polytope) *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then the dimension of* $\operatorname{conv} S$ *is* $(N+1)(T-1)$*, that is,* $\operatorname{conv} S$ *is full-dimensional.* □

Proof First note that since $S \subseteq \mathbb{R}^{(N+1)(T-1)}$ it holds that $\dim(\operatorname{conv} S) \leq (N+1)(T-1)$. Let the vectors $(x^k, z^k) \in \mathbb{B}^{(N+1)(T-1)}$, $k \in \{0, \ldots, (N+1)(T-1)\}$, be given by the following. For $i \in \mathcal{N}$ and $t = 1, \ldots, T-1$, let $x^k_{it} = 0$ if $k \in \{(N+1)(t-1)+i, (N+1)t\}$ and $x^k_{it} = 1$ otherwise. For $t = 1, \ldots, T-1$, let $z^k_t = 0$ if $k = (N+1)t$ and $z^k_t = 1$ otherwise. Since $T_i \geq 2$ for $i \in \mathcal{N}$ it holds that $\sum_{t=\ell}^{T_i+\ell-1} x^k_{it} \geq 1$ for all $i \in \mathcal{N}$, all $\ell \in \{1, \ldots, T - T_i\}$, and all $k \in \{0, \ldots, (N+1)(T-1)\}$.

Moreover, for all $t = 1, \ldots, T-1$ and $k \in \{0, \ldots, (N+1)(T-1)\}$ such that $z^k_t = 0$ it holds that $x^k_{it} = 0$, $i \in \mathcal{N}$. It follows that $(x^k, z^k) \in S$, as defined in (4). Further, it can be verified that the only solution to the system

$$\sum_{k=0}^{(N+1)(T-1)} x^k_{it}\alpha_k = 0, \quad i \in \mathcal{N}, \qquad \sum_{k=0}^{(N+1)(T-1)} z^k_t \alpha_k = 0, \quad t = 1, \ldots, T-1, \qquad \sum_{k=0}^{(N+1)(T-1)} \alpha_k = 0,$$

is $\alpha_k = 0$, $k \in \{0, \ldots, (N+1)(T-1)\}$, implying that the vectors (x^k, z^k), $k \in \{0, \ldots, (N+1)(T-1)\}$, are affinely independent. Hence, it holds that $\dim(\operatorname{conv} S) \geq (N+1)(T-1)$, thus implying that $\dim(\operatorname{conv} S) = (N+1)(T-1)$. The proposition follows. □

The replacement polytope is *not* full-dimensional if $T_i = 1$ for some $i \in \mathcal{N}$, since it then holds that $x_{it} = z_t = 1$, $t = 1, \ldots, T-1$, for all $(x,z) \in \operatorname{conv} S$. Letting $A^{=}$ denote the matrix corresponding to the equality subsystem of $\operatorname{conv} S$, this would yield that $\operatorname{rank} A^{=} \geq 2(T-1)$ and thus that $\dim(\operatorname{conv} S) \leq (N-1)(T-1)$. 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.

PROPOSITION 6 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities $\sum_{t=\ell}^{T_i+\ell-1} x_{it} \geq 1$, $\ell = 1, \ldots, T - T_i$, $i \in \mathcal{N}$, defines a facet of* $\operatorname{conv} S$. □

PROOF Since $T_i \geq 2$ for $i \in \mathcal{N}$, $\operatorname{conv} S$ is full-dimensional (cf. Proposition 5). Hence, we can use the uniqueness characterization of the facet description from Proposition 3 to show the assertion.

For each $r \in \mathcal{N}$ and each $\ell \in \{1, \ldots, T - T_r\}$, let $\widehat{F}_{r\ell} = \{\, (x,z) \in \operatorname{conv} S \mid \sum_{t=\ell}^{T_r+\ell-1} x_{rt} = 1 \,\}$. Further, let $x_{it}^0 = z_t^0 = 1$, $i \in \mathcal{N}$, $t \in \{1, \ldots, T-1\}$. Since $T_i \geq 2$ it follows that $(x^0, z^0) \in S \setminus \widehat{F}_{r\ell}$. Then, defining the vector $(x^{\mathrm{A}}, z^{\mathrm{A}})$ as $x_{it}^{\mathrm{A}} = 0$ if $i = r$ and $t \in \{\ell+1, \ldots, T_r + \ell - 1\}$, $x_{it}^{\mathrm{A}} = 1$ otherwise, and $z_t^{\mathrm{A}} = 1$, $t \in \{1, \ldots, T-1\}$, it follows that $(x^{\mathrm{A}}, z^{\mathrm{A}}) \in \widehat{F}_{r\ell}$ and hence that $\widehat{F}_{r\ell}$ is a proper face of $\operatorname{conv} S$. Moreover, there exist values of $\lambda \in \mathbb{R}^{N \times (T-1)}$, $\mu \in \mathbb{R}^{T-1}$, and $\rho \in \mathbb{R}$ such that the equation

$$\sum_{t=1}^{T-1} \left(\sum_{i \in \mathcal{N}} \lambda_{it} x_{it} + \mu_t z_t \right) = \rho \tag{7}$$

is satisfied for all $(x,z) \in \widehat{F}_{r\ell}$. We will show that for any value of $\alpha \in \mathbb{R}$, in a solution to (7) it holds that $\lambda_{it} = \alpha$ if $i = r$ and $t \in \{\ell, \ldots, T_r + \ell - 1\}$, $\lambda_{it} = 0$ otherwise, $\mu_t = 0$, $t \in \{1, \ldots, T-1\}$, and $\rho = \alpha$.

For each $i \in \mathcal{N} \setminus \{r\}$ and each $t \in \{1, \ldots, T-1\}$, let, for $j \in \mathcal{N}$ and $k \in \{1, \ldots, T-1\}$, $x_{jk}^1 = 0$ if $j = i$ and $k = t$, $x_{jk}^1 = x_{jk}^{\mathrm{A}}$ otherwise, and $z_t^1 = z_t^{\mathrm{A}}$, $t \in \{1, \ldots, T-1\}$. It follows that $(x^1, z^1) \in \widehat{F}_{r\ell}$. The vectors $(x^{\mathrm{A}}, z^{\mathrm{A}})$ and (x^1, z^1), respectively, inserted in (7) then yield that $\lambda_{it} = 0$ for all $i \in \mathcal{N} \setminus \{r\}$ and all $t \in \{1, \ldots, T-1\}$.

For each[1] $t \in \{1, \ldots, \ell - 1\} \cup \{T_r + \ell + 1, \ldots, T-1\}$, let, for $i \in \mathcal{N}$ and $k \in \{1, \ldots, T-1\}$, $x_{ik}^2 = 0$ if $i = r$ and $k = t$, $x_{ik}^2 = x_{ik}^{\mathrm{A}}$ otherwise, and let $z_t^1 = z_t^{\mathrm{A}}$, $t \in \{1, \ldots, T-1\}$. It follows that $(x^2, z^2) \in \widehat{F}_{r\ell}$. The vectors $(x^{\mathrm{A}}, z^{\mathrm{A}})$ and (x^2, z^2), respectively, inserted in (7) then yield that $\lambda_{rt} = 0$ for all $t \in \{1, \ldots, \ell-1\} \cup \{T_r + \ell + 1, \ldots, T-1\}$.

Further, let, for $i \in \mathcal{N}$, $x_{it}^{\mathrm{B}} = 0$ if $i = r$ and $t = \ell$, $x_{it}^{\mathrm{B}} = 1$ if $i = r$ and $t = T_r + \ell - 1$, $x_{it}^{\mathrm{B}} = x_{it}^{\mathrm{A}}$ otherwise, and let $z_t^{\mathrm{B}} = z_t^{\mathrm{A}}$, $t \in \{1, \ldots, T-1\}$. Moreover, let, for $i \in \mathcal{N}$, $x_{it}^3 = 0$ if $i = r$ and $t = T_r + \ell$, $x_{it}^3 = x_{it}^{\mathrm{B}}$ otherwise, and let $z_t^3 = z_t^{\mathrm{B}}$, $t \in \{1, \ldots, T-1\}$. It follows that $(x^3, z^3) \in \widehat{F}_{r\ell}$. The vectors $(x^{\mathrm{B}}, z^{\mathrm{B}})$ and (x^3, z^3), respectively, inserted in (7) then yield that $\lambda_{r, T_r+\ell} = 0$. The equation (7) can then be rewritten as

$$\sum_{t=1}^{T-1} \mu_t z_t + \sum_{t=\ell}^{T_r+\ell-1} \lambda_{rt} x_{rt} = \rho \tag{8}$$

For each $t \in \{1, \ldots, T-1\} \setminus \{\ell, T_r + \ell\}$, let, for $i \in \mathcal{N}$, $x_{ik}^4 = 0$ if $k = t$, $x_{ik}^4 = x_{ik}^{\mathrm{A}}$ otherwise, and let $z_k^4 = 0$ if $k = t$, and $z_k^4 = z_k^{\mathrm{A}}$ otherwise. It follows that $(x^4, z^4) \in \widehat{F}_{r\ell}$. The vectors $(x^{\mathrm{A}}, z^{\mathrm{A}})$ and (x^4, z^4), respectively, inserted in (8) then yield that $\mu_t = 0$ for all $t \in \{1, \ldots, T-1\} \setminus \{\ell, T_r + \ell\}$.

Further, for each $t \in \{\ell, T_r + \ell\}$, let, for $i \in \mathcal{N}$, $x_{ik}^5 = 0$ if $k = t$, $x_{ik}^5 = x_{ik}^{\mathrm{B}}$ otherwise, and let $z_k^5 = 0$ if $k = t$, and $z_k^5 = z_k^{\mathrm{B}}$ otherwise. It follows that $(x^5, z^5) \in \widehat{F}_{r\ell}$. The vectors $(x^{\mathrm{B}}, z^{\mathrm{B}})$ and (x^5, z^5), respectively, inserted in (8) then yield that $\mu_\ell = \mu_{T_r+\ell} = 0$. Equation (8) can then be rewritten as

$$\sum_{t=\ell}^{T_r+\ell-1} \lambda_{rt} x_{rt} = \rho \tag{9}$$

For each $t \in \{\ell+1, \ldots, T_r + \ell - 1\}$, let for $i \in \mathcal{N}$ and $k \in \{1, \ldots, T-1\}$, $x_{ik}^6 = 0$ if $i = r$ and $k = \ell$, $x_{ik}^6 = 1$ if $i = r$ and $k = t$, and $x_{ik}^6 = x_{ik}^{\mathrm{A}}$ otherwise, and let $z_t^6 = z_t^{\mathrm{A}}$, $t \in \{1, \ldots, T-1\}$. It follows that $(x^6, z^6) \in \widehat{F}_{r\ell}$. The vectors $(x^{\mathrm{A}}, z^{\mathrm{A}})$ and (x^6, z^6), respectively, inserted in (9) then yield that $\lambda_{r\ell} = \lambda_{rt}$. Hence, λ_{rt} is constant over $t \in \{\ell, \ldots, T_r + \ell - 1\}$ and we define $\lambda_{rt} = \lambda$, $t \in \{\ell, \ldots, T_r + \ell - 1\}$. Since

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

$(x^{\mathrm{A}}, z^{\mathrm{A}}) \in \widehat{F}_{r\ell}$ it follows that $\lambda = \rho$. Letting $\alpha = \rho$, the equation (9) can be written as $\sum_{t=\ell}^{T_r+\ell-1} \alpha x_{rt} = \alpha$. Proposition 3 then yields that the inequality $\sum_{t=\ell}^{T_r+\ell-1} x_{rt} \geq 1$ defines a facet of $\operatorname{conv} S$. □

The technique used to prove Proposition 6 can be applied to Propositions 7–9, whose proofs are therefore omitted here.

PROPOSITION 7 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities $x_{it} \leq z_t$, $i \in \mathcal{N}$, $t = 1, \ldots, T-1$, defines a facet of* $\operatorname{conv} S$. □

PROPOSITION 8 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities $z_t \leq 1$, $t = 1, \ldots, T-1$, defines a facet of* $\operatorname{conv} S$. □

PROPOSITION 9 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities $x_{kt} \geq 0$, $k \in \mathcal{N} : T_k \geq 3$, $t = 1, \ldots, T-1$, defines a facet of* $\operatorname{conv} S$. □

The inequalities in Proposition 9 do *not* define facets for any $k \in \mathcal{N}$ such that $T_k \leq 2$ due to the following. If $T_k = 2$ then, for each $s \in \{1, \ldots, T-2\}$, $x_{ks} = 0$ 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 $\operatorname{rank} A^{=} \geq 2$, where $A^{=}$ denotes the matrix corresponding to the equality subsystem of $\operatorname{conv} S$. Letting $\tilde{F}_{ks} = \{ (x,z) \in \operatorname{conv} S \mid x_{ks} = 0 \}$, it follows that $\dim \tilde{F}_{ks} \leq (N+1)(T-1) - 2$, which implies that $\tilde{F}_{ks}$ is *not* a facet of $\operatorname{conv} S$.

Now, the set S is defined by the constraints (1), (2), and

$$x_{it} \geq 0, \quad z_t \leq 1, \quad x_{it},\ z_t \in \mathbb{B}, \qquad t = 1, \ldots, T-1, \quad i \in \mathcal{N}, \tag{10}$$

and it follows from Propositions 6–9 that all of the inequalities necessary in the description of the set S define facets of $\operatorname{conv} S$. A natural question then arises: Is $\operatorname{conv} S$ completely described by the continuous relaxation of the system (1)–(2), (10)? Unfortunately, this is not the case, which is shown by the following example.

EXAMPLE 10 (continuous relaxation) Consider a system with $N = 2$, $T_1 = 3$, $T_2 = 4$, and $T = 5$. Then the problem to

$$\begin{array}{ll} \text{minimize} & x_{11} + x_{12} + 2x_{13} + x_{14} + x_{21} + 100x_{22} + 100x_{23} + x_{24} + 10z_1 + 10z_2 + z_3 + 10z_4, \\ \text{subject to} & \text{(1)–(2) and (10)}, \end{array}$$

has the optimal solution

$$(x_{11}, x_{12}, x_{13}, x_{14}; x_{21}, x_{22}, x_{23}, x_{24}; z_1, z_2, z_3, z_4) = (0,0,1,0;1,0,0,0;1,0,1,0), \tag{11}$$

with objective function value 14. Relaxing the integrality requirements, yields the optimal solution

$$(x_{11}, x_{12}, x_{13}, x_{14}; x_{21}, x_{22}, x_{23}, x_{24}; z_1, z_2, z_3, z_4) = (0.5,0,0.5,0.5;0.5,0,0,0.5;0.5,0,0.5,0.5), \tag{12}$$

with objective function value 13.5. Hence the convex hull of feasible solutions to the system (1)–(2), (10) is not completely defined by the inequalities therein. □

Example 10 shows that the inequalities in (1)–(2), (10) are not sufficient to describe $\operatorname{conv} S$. However, according to the Propositions 6–9, almost all of these inequalities define facets of $\operatorname{conv} S$. Since, by Proposition 5, $\operatorname{conv} S$ is full-dimensional (under reasonable assumptions) the minimal description of $\operatorname{conv} S$ is unique. Therefore, all of these facets are necessary in the description of $\operatorname{conv} S$.

To completely describe $\operatorname{conv} S$ we need however also facets other than those in (1)–(2), (10). By using Chvátal–Gomory rounding (see [15, p. 210]) it can be shown that the inequalities

$$z_\ell + \sum_{t=\ell+1}^{\ell+T_i-2} (x_{it} + x_{jt}) + z_{\ell+T_i-1} \geq 2, \quad \ell \in \{1, \ldots, T-T_i\}, \quad i,j \in \mathcal{N} \colon 2 \leq T_j \leq T_i - 1 \leq 2(T_j - 1), \tag{13}$$

are valid for S. For the replacement problem in Example 10, the inequalities (13) reduce to the single inequality $z_1 + x_{12} + x_{13} + x_{22} + x_{23} + z_4 \geq 2$, which is *not* satisfied by the optimal solution (12) to the continuous relaxation of the replacement problem in Example 10. In fact, by adding this inequality to the continuous relaxation, the optimal solution given in (11) is obtained. Using the same technique as in the proof of Proposition 6 the following result can be shown:

PROPOSITION 11 *If $T_i \geq 2$ for all $i \in \mathcal{N}$, then each of the inequalities* (13) *defines a facet of* $\operatorname{conv} S$. □

The Chvátal–Gomory rounding applied here can be generalized to find new classes of facets for the general replacement problem. However, it still remains to investigate the strength of the continuous relaxation resulting from new classes of facets being added to the replacement problem.

4 Towards a model for the stochastic optimization problem

We introduce stochastic modeling of the lifes of the components, to obtain a more applicable mathematical model. We also treat a more informative data type when one has measurements on the path to failure of a component. A further topic is discretization of the obtained distributions, which is necessary for the optimization problem.

We treat the distribution of time to failure for only one component; in a real type situation one typically has measurements on failures for several components. If an aircraft engine is made up of m components and $T_1, \ldots, T_m$ are their failure times (lives), then the aircraft engine's life is $T = f(T_1, \ldots, T_m)$ for some function f. If f is known the approach below can be applied to estimation of the distribution of T. If the individual components are dependent one needs to model and estimate the simultaneous distribution of $T_1, \ldots, T_m$ and the approach below needs to be modified.

Let us treat a fixed *component* for which we have measurements on several engines, or *units*. Assume that $\{x(t), t \in [0,T]\}$ is an $\mathbb{R}^p$-valued stochastic process modeling the crack size for the component, during it's life span $[0,T]$. When we have n units the $\mathbb{R}^p$ valued processes $x_1(t), \ldots, x_n(t)$ model the crack sizes in the units, and thus $\{x_i(t)\}_{i=1}^n$ is a sequence of independent and identically distributed stochastic processes. During it's life the unit i will be serviced, at some time points $t_{i1}, t_{i2}, \ldots$, defined as

$$t_{i1} = \inf\{t \geq 0 : x_i(t) \in C\}, \quad t_{i,k+1} = \inf\{t > t_{ik} : x_i(t) \in C\} - t_{ik}, \text{ for } k \geq 1, \tag{14}$$

where $C \subseteq \mathbb{R}^p$ is the critical region for the process.

We assume that C is independent of t, i.e. that the time dynamics of the process do not influence the reliability of the choice to make a repair. Furthermore C is assumed to be the same for all units i, so that the unit's failure times $t_i = (t_{i1}, \ldots, t_{i,ni})$ are independent random vectors.

The objective of Section 4.1 is the transfer of structural information: There are two types of engines, I and II. They are structurally different, and the amount of data differs significantly; for type I there exists a large data set whereas for type II data is limited. The more interesting engine is type II, being a replacement of type I. Let $x(t)$ be a random element (a real number, $p = 1$, or a finite-dimensional vector, $p > 1$) with unknown distribution $F = F(;\theta)$ parameterized by $\theta \in \Theta \subset \mathbb{R}^s$, with $s < \infty$. Let F_I (resp. F_{II}) denote the distribution for type I (resp. II) engines. A natural assumption is that the distribution functions belong to the same parametric class of distributions $\mathcal{P} = \{F(\cdot;\theta) : \theta \in \Theta\}$, and that only the parameter values differ: $F_I = F(\cdot;\theta_I)$ and $F_{II} = F(\cdot;\theta_{II})$, with $F(\cdot,\cdot)$ a known function.

The observations of the feature processes are of two types. One type of data consists of times to service, t_{ik}, for the units, with or without the corresponding feature values, $x_i(t_{ik})$. A finer type of data consists of repeated measurements in the same cycle k of the feature process $x_i(s_{i1}), x_i(s_{i2}), \ldots$, with time points s_{ij} possibly passing the time to repair so that possibly $s_{ij} > t_{ik}$ for some j and k. Thus, assume that the time-dependent crack size $x_i(t)$ in unit i of a component has distribution $F_{x_i} = F_x(\cdot;\theta_x)$ with unknown parameter θ_x. The stochastic process $x_i(t)$ is not completely observed, because of a measurement error ϵ_i with distribution $F_\epsilon(\cdot;\theta_\epsilon)$ at times s_{ik}: the observations consist of $g(x(s_{ik}), \epsilon_i)$, for some function g, and are distribution according to $F(\cdot;\theta_x,\theta_\epsilon)$. Let C be the critical region for $x_i(t)$ and $t_{ik} = \inf\{t : x_i(t) \in C\}$ the failure time. It is then possible to obtain the distribution $F_{t_{ik}} = F_{t_{ik}}(\cdot,\theta_x,\theta_\epsilon)$ of t_{ik}, treating the parameter θ_ϵ as a nuisance parameter, and t_{ik} as the interesting random variable for which we want to estimate the distribution based on the data $g(x(s_{ik}), \epsilon_i)$.

The resulting estimators of life distributions are continuous. Since the optimization models in Sections 2 and 4.2 treat discrete probability functions we discretize the distributions, using as few points of support as possible, in order to limit the complexity of the corresponding optimization model; this is the topic for Section 4.2.2. In Section 4.3 we then draw some conclusions and discuss the implications of the stochastic modeling for the optimization model.

4.1 Structural models for time between repairs

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

The data consists of times between repairs for a number of units for three components in the flame holder. For a fixed component and every unit i we observe a sequence $t_i = \{t_{i1}, ..., t_{i\ell}\}$ of times between repairs where observation ℓ is possibly (right-) censored, meaning that the time to repair $\ell + 1$ is longer than the time observed. We model different details as independent, that is, the corresponding random vectors t_i and t_j are independent if $i \neq j$.

Consider a fixed unit i for an arbitrary component. Let t_{ij} be the time between the $j-1$'th and the j'th repair, and let $F_j(t) = P(t_{ij} \leq t)$ be the corresponding distribution function. Let $N_i(t) = \#\{t_{ij} \leq t \mid j = 1, \ldots, \ell_i\}$ be the corresponding counting process that counts the number of events (repairs) that have occurred by time t.

DEFINITION 12 (renewal process) *An independently distributed 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 13 (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-homogeneous Poisson process (NHPP); if $w(t) \equiv w$ is a constant function the process is a homogeneous Poisson process.* □

4.1.1 Survival analysis

For a particular unit, 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 time observed. One typically assumes that censoring is due to other mechanisms than the ones governing the failure time, cf. Andersen et al. [2]. Thus the data consists of pairs (t_{ij}, δ_{ij}) with t_{ij} being the observed j'th time between repairs for unit i and $\delta_{ij} \in \{0, 1\}$ indicating whether the observed time is a failure or a censoring. Let us suppress the index j in the sequel.

Define the survival function $S_i(t) = 1 - F_i(t)$ and the hazard function $h_i(t) = -S_i'(t)/S_i(t)$. Given (possibly right censored) data $\{(t_i, \delta_i)\}_{i=1}^n$, the standard estimator of the survival function is the Kaplan–Meier estimator $S_n(t) = \prod_{i:t_i \leq t}\left(1 - \frac{\delta_i}{Y_n(t_i-)}\right)$, with $t_i- = \lim_{h\downarrow 0} t_i - h$, where $Y_n(t) = \sum_{i=1}^n 1\{t_i > t\}$ is the number of units at risk to fail at time t. The standard estimator of integrated hazard function $H(t) = \int_0^t h(u)\,du$ is the Nelson–Aalen estimator $H_n(t) = \sum_{t_i \leq t} \frac{\delta_i}{Y_n(t_i-)}$.

A refinement of the above is possible if there are variables that influence the distribution of the time between failures. Thus we assume that for each unit i we are able to measure a vector $z_i = (z_{i1}, \ldots, z_{ip})$ of covariates and that these affect the survival and hazard functions so that $S_i(t) = S(t; z_i)$ and $h_i(t) = h(t; z_i)$ are unit-dependent survival and hazard functions. The standard approach to modeling for inclusion of covariates is the Cox proportional hazards model, cf. Cox [8] and Andersen et al. [2]. 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 e.g. a Cox regression model, since the latter has no physical justification. 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 a formal expression for the distribution of the time between repairs.

4.1.2 Model fit

Let $\{t_i\}_{i=1}^n$ be times between repairs for n units with unknown distribution F. We make a formal test of the hypothesis $H : F = F_0$, using the test statistic $M = \frac{1}{n}\sum_{i=1}^n (E_{F_0}(T_i) - t_i)^2$, where T_i is the random variable of which t_i is an observation and $E_F(T_i)$ is the expectation of T_i under the null distribution F_0.

For the NSRP model T_i is Weibull distributed and $E_{F_0}[T_i]$ then has a known parametric form, cf. Table 1. The fact that M_e is mostly bigger than M may be a consequence of the true distribution having shorter tails than the Weibull distribution.

Table 1: *Prediction errors, observed (M) and expected ($M_e = E_{F_0}(T_1 - E_{F_0}(T_1))^2$), of the NSRP model for components 1, 2, and 3 for the four first failures and average time to failure.*

Component		Failure 1		Failure 2		Failure 3		Failure 4		Average	
	$10^4\times$	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
3		7.20	7.45	0.22	0.21	0.29	0.26	0.28	0.26	1.05	1.08

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

$$E_{F_0}[T_i] = \int_0^\infty e^{-[H(t_0+t)-H(t_0)]}dt,$$

where H is the cumulative intensity (or cumulative hazard function) $H(t) = \int_0^t h(u)\,du$ and where $h = w$ is defined in Definition 13. It is not possible to estimate H for values of t larger than the largest observed life without assuming parametric models; one remedy for calculating $E_{F_0}(T)$ is to integrate up to the largest observation, cf. Table 2.

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

Component		Failure 1	Failure 2	Failure 3	Failure 4	Average
	$10^4\times$	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; note that M_{NHPP} and M_{NSRP} are the average errors in the respective models given in Tables 1 and 2.

Table 3: *Relative errors for the two models NSRP and NHPP (for the average time to failure).*

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

Repairs are performed at repair stations within a close range of the aircraft (A) and the main central repair station (B). To test for a difference between the two types of repairs, we estimate the mean time to repair for the three details but distinguish between the repairs A and B. In Figure 3 this is shown for the first five repairs with 95% confidence interval for the mean. There seems to be a difference between the repairs A and B, at least for components of type 1 and 3.

To test whether F_n depends on n, we have estimated θ, α, and the mean μ_w under the Weibull distribution, and of the mean μ under the the non-parametric distribution, for the time to first failure and for the following five times to failure after repair for components 1, 2, and 3, cf. Svensson [17] for more details. We noticed 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, $i \geq 2$. There seems to be two classes of repairs: repair of new components and repair of old ones.

To test whether the time to the next repair decreases with the number of repairs, we use the model

$$F_{T_n}(t) = 1 - e^{-(\frac{1}{\theta p^n}t)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0, p > 0), \tag{15}$$

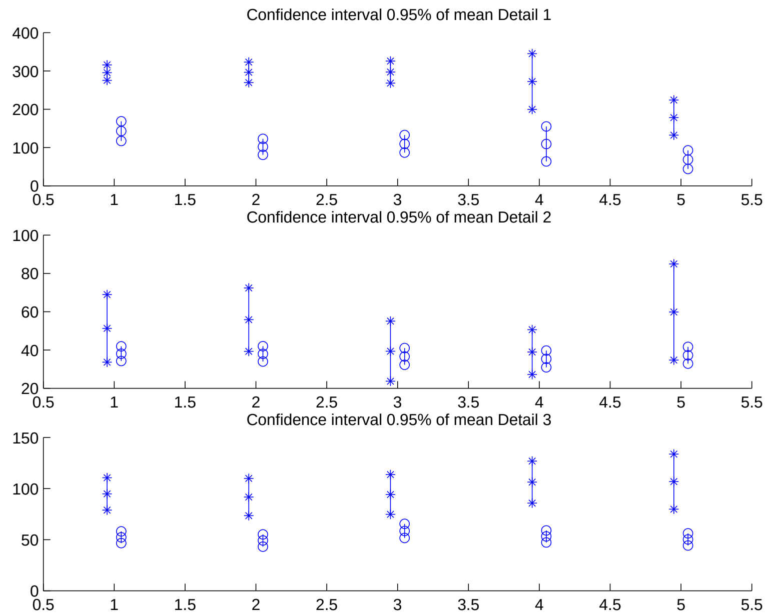

Figure 3: *Confidence intervals for the expected time to failure after repair numbers 1 to 5. Rings and stars represent repair type A and B, respectively.*

where n is the repair number. Then the expected time to failure after repair n is $E[T_n] = \theta p^n \cdot \Gamma(\frac{1}{\alpha} + 1)$, and $p < 1$ indicates aging. Maximum likelihood estimates of the parameters (θ, α, p) are shown in Table 4 together with 95% confidence intervals 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 4: *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 Crack length modeling

Let $\{x(t) : t \geq 0\}$ be a real valued stochastic process that describes the growth of a crack in a part 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 crack length $a_i(t)$ of crack i at time t is modeled by

$$a_i(t) = \begin{cases} a_0, & \text{if } t < S_i, \\ a_0 + C_i \cdot (t - S_i)^b, & \text{if } t \geq S_i, \end{cases} \qquad i \in \{1, \dots, k\}, \tag{16}$$

where C_i and S_i are stochastic variables and $b > 0$ and a_0 are parameters, cf. Svensson [18].

The cracks are not observed directly: For each crack $i \in \{1, \dots, n\}$ there are n_i observations $x_{ij} = x_i(t_{ij})$ at times $t_{i1} < \dots < t_{in_i}$, following the model

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

with ε_{ij} being a sequence of independent $N(0, \sigma_\varepsilon^2)$ distributed random variables. This implies that the distribution of $X_i(t_{ij})$ is, conditionally on (c_i, s_i), a mixture of a discrete and a continuous distribution, the discrete random variable having a point mass a_0 at s_i.

Let $\theta = (\psi, \sigma_\varepsilon^2, b)$ be the parameter vector; here ψ are the parameters determining the distribution for the pairs of random variables (C_i, S_i) in (16). For crack i we want to find the the distribution of the time $T_i = \inf\{t : a_i(t) \geq a_{\max}\}$, where $a_{\max}$ denotes the minimum length that is registered as a crack.

Assume that the previous observations $\mathbf{x}_i = (x_i(t_{i1}), \ldots, x_i(t_{in_i}))$ of crack i are given and that the parameters θ of the model are known. Then, using model (16), it is possible to derive expressions for the conditional distribution

$$P(T_i \leq t \mid \mathbf{X}_i = \mathbf{x}_i) \tag{17}$$

and the corresponding density function $f_{T_i|\mathbf{X}_i}(t|\mathbf{x}_i;\theta)$. This implies that $f_{(C_i,S_i)|\mathbf{X}_i}(c_i, s_i \mid \mathbf{x}_i;\theta) = f_{\mathbf{X}_i|C_i,S_i}(\mathbf{x}_i \mid c_i, s_i; b, \sigma_\varepsilon^2) \cdot f_{C_i,S_i}(c_i, s_i;\psi)/f_{\mathbf{X}_i}(\mathbf{x}_i;\theta)$, via Bayes' formula. Using the model descriptions it is possible to obtain expressions for all factors in this conditional density.

Let A_i be the crack length of crack i at a fixed time t. Similarly, we obtain $f_{A_i|\mathbf{X}_i}(a \mid \mathbf{x}_i;\theta) = f_{\mathbf{X}_i|A_i}(\mathbf{x}_i \mid a;\theta) \cdot f_{A_i}(a;\theta) f_{\mathbf{X}_i}(\mathbf{x}_i;\theta)$, and it is possible to obtain expressions for all factors in this conditional density. The likelihood for the parameters given data of cracks $\underline{\mathbf{x}} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is

$$L(\theta) = f_{\underline{\mathbf{X}}}(\underline{\mathbf{x}};\theta) = \prod_{i=1}^{n} f_{\mathbf{X}_i}(\mathbf{x}_i;\theta), \tag{18}$$

In order to take the parameter uncertainty into account we use a profile likelihood approach, cf. Pawitan [16], and define the predictive profile likelihood for T_i given $\mathbf{x}_i$ as

$$\tilde{L}(t \mid \mathbf{x}_i; \underline{\mathbf{x}}_{-i}) = \sup_\theta \; f_{T_i|\mathbf{X}_i}(t \mid \mathbf{x}_i;\theta) \; f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i};\theta). \tag{19}$$

where $\underline{\mathbf{x}}_{-i} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_k, \}$.

If the number n of cracks is large the predictive profile likelihood in (19) 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 $L(\hat{\theta}, t \mid \mathbf{x}_i; \underline{\mathbf{x}}_{-i}) = f_{T_i|\mathbf{X}_i}(t \mid \mathbf{x}_i;\hat{\theta}) \cdot f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i};\hat{\theta})$. The predictive profile likelihood for inference on the crack length at a specific time is then given by $\tilde{L}(A_i \mid \mathbf{x}_i; \underline{\mathbf{x}}_{-i}) = \sup_\theta(f_{A_i|\mathbf{X}_i}(a \mid \mathbf{x}_i;\theta) \cdot f_{\underline{\mathbf{X}}_{-i}}(\underline{\mathbf{x}}_{-i};\theta))$.

4.1.4 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_i and S_i, 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 so called pri-engines, that are used extensively and 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_i and S_i. Figure 4 illustrates the cracks and the crack model (16) fitted to the cracks with a least squares method. From each picture we get an observation of S_i (censored if no crack was detected) and an observation of C_i if a crack was detected.

Using a similar procedure for all cracks (more than the four in Figure 4) indicates that C_i and S_i are uncorrelated, which makes it feasible to assume independence between C_i and S_i. Furthermore, by examining the empirical distribution of C_i and S_i we find that the log normal distribution gives a reasonable fit; we therefore assume that both C_i and S_i are log-normally distributed with parameters (μ_s, σ_s) and (μ_c, σ_c), respectively.

4.1.5 Model illustration

Assume that we have observed the cracks on the first three components and want to predict when the crack on the fourth one reaches the length $a_{\max} = 30$ mm, that is, to find the distribution of $T = \inf\{t : a(t) \geq a_{\max}\}$ given the observations of component four. It is also possible to update the distribution of T when we get new observations, at 200, 400, 600, and 800 flight hours (FH), respectively.

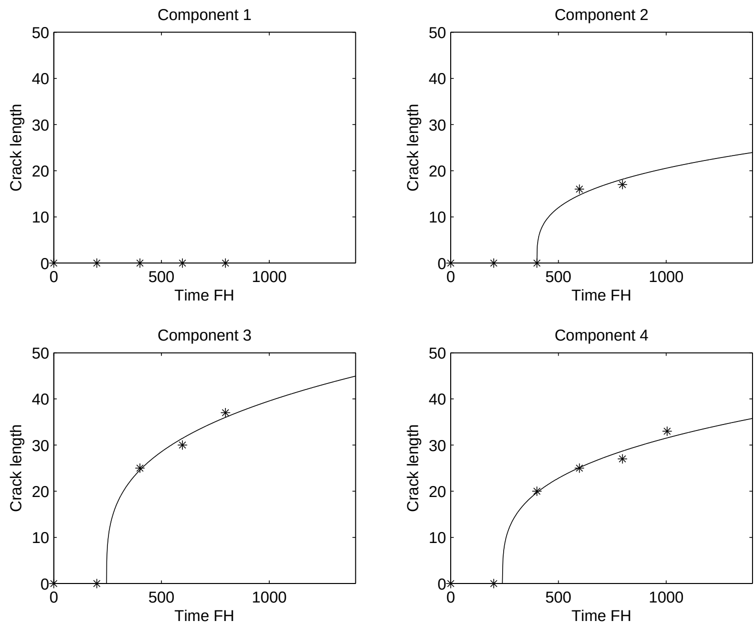

Figure 4: *Model* (16) *with* $a_0 = 0$ *fitted to the four cracks.*

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 (18), i.e. $\psi = (\mu_s, \sigma_s, \mu_c, \sigma_c)$. First we calculate the distribution of T using (17), ignoring the uncertainty in the parameter estimation. The solid line in Figure 5 illustrates the distribution of T when ignoring parameter estimation uncertainty and $\theta = \hat{\theta}$.

Using the profile likelihood approach (19) we see how much the uncertainty in the parameter estimates affects the results. In Figure 5 the lines with stars are plots 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 6 shows the crack growth over time given the information in our observations. Integrating the function in Figure 6 with respect to crack length yields a marginal function of value 1 for all points in time. 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 there is 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. The other pictures in Figure 6 indicate that even if we are fairly sure of the crack length at a fixed time the distribution of when the crack reaches a specific length will have a large variance.

Combining the probability model with the profile likelihood based approach yields a comparison of the effect of the uncertainty in the parameters on the distribution of the crack length at a fixed time. This is plotted in Figure 7 for the time points 500 FH and 1000 FH.

4.2 Optimal discretization of a continuous distribution

We develop the case with one stochastic component; it can be extended to several stochastic components. The life U of a new stochastic component is modeled with a distribution $\breve{G}$ and the remaining life of a functioning stochastic component of age of u_0 is modeled with a distribution defined by $G(u) = \frac{\breve{G}(u+u_0)-\breve{G}(u_0)}{1-\breve{G}(u_0)}$. We assume that $\breve{G}'(u) > 0$ if $u > 0$ and that U is a non-negative random variable.

We make two simplifications: First, in the sequence of life distributions for the stochastic component only the first life distribution is modeled as a random variable, the remaining ones are replaced by a functional of the distribution such as the expected value or the median. Second, we need to discretize the life distribution, since the optimization model is defined using discrete times.

We distinguish between the first and the second stage models. The first takes into consideration all

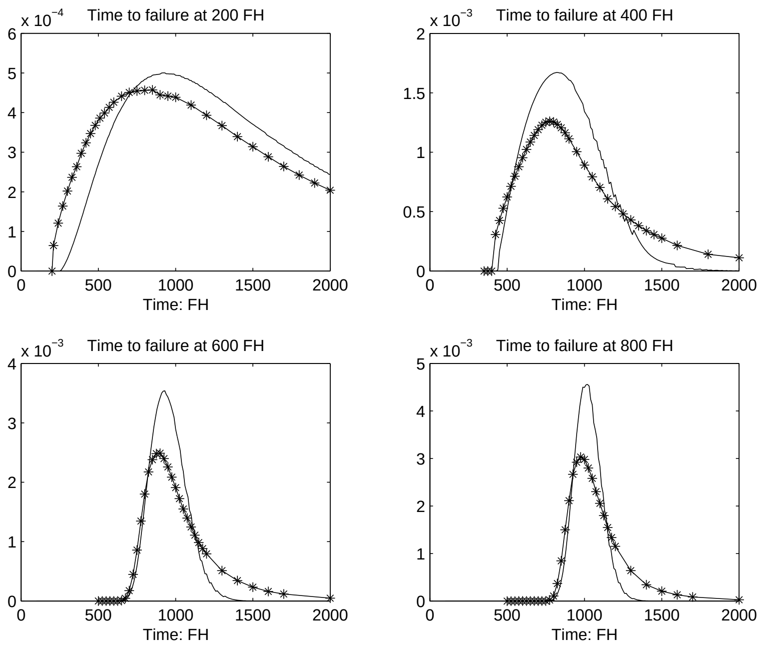

Figure 5: *Distribution of the time* $T = \inf\{t : a(t) \geq a_{\max}\}$ *when the crack of component four will reach* $a_{\max} = 30$ *mm both considering uncertainty in parameter estimation (line with stars) and without uncertainty (solid line). The distribution is updated with the observations at 200 FH (upper left), 400 FH (upper right), 600 FH (lower left) and 800 FH (lower right).*

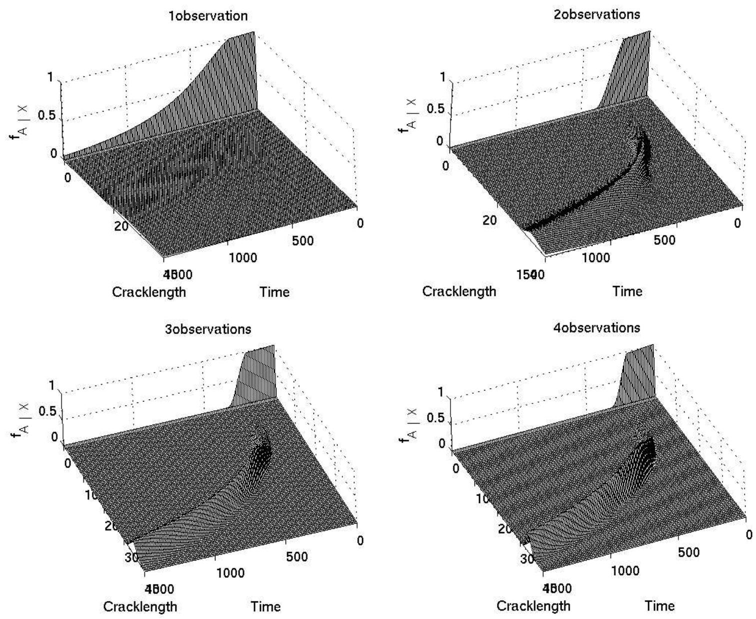

Figure 6: *A three dimensional illustration of how the distribution of the crack length at a fixed time of component four varies with the numbers of observations. The distribution is updated each time a new observation is made: first observation at 200 FH (upper left), second observation at 400 FH (upper right), third observation at 600 FH (lower left) and fourth observation at 800 FH (lower right). The values at the crack length* 0 *corresponds to the probability (a point mass) that there are no cracks.*

possible future events, while the second stage model contains one model for each future event.

We minimize the expected cost of maintaining the engine during a fixed time period containing T

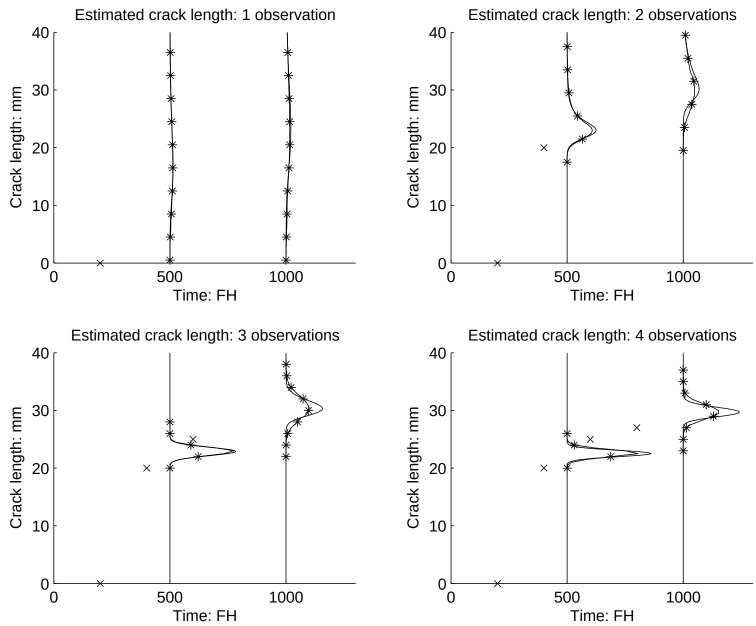

Figure 7: *Distribution of the crack length at the times 500 and 1000 FH for component four both considering uncertainty in parameter estimation (line with stars) and without uncertainty (solid line). The distributions are updated with the observations at 200 flight hours (upper left), 400 flight hours (upper right), 600 flight hours (lower left) and 800 flight hours (lower right). The observations are marked in the pictures with "x".*

equidistant time points, with components allowed to be replaced only 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 $\tilde{\tau}(u) \in \{1, \ldots, T-1\}$. The life of each replacing stochastic component is defined as $\tau \approx E_{\breve{G}} U$ (see Altenstedt [1]) such that $\tau \in \{1, \ldots, T-2\}$ (if $\tau \geq T-1$ the stochastic component is replaced at most once during the time horizon considered). In addition to the costs defined in Section 2, c denotes the cost for replacing a 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. The binary variables s_t are defined as

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

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. Let $\mathbf{x}_1 = (x_{11}, \ldots, x_{N1}, s_1, z_1)$ be the replacement strategy vector $\hat{\mathbf{x}}_1 \in \operatorname{argmin}_{\mathbf{x}_1 \in \mathbb{B}^{N+2}} F(\mathbf{x}_1)$, the optimal replacement strategy for the first stage, where

$$F(\mathbf{x}_1) = \int_0^\infty f(\mathbf{x}_1, u)\, dG(u) = E_G[f(\mathbf{x}_1, U)]. \tag{20}$$

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

$$f(\overline{\mathbf{x}}_1, \tilde{\tau}(u)) = \min \qquad \sum_{t=1}^{T-1} \left(\sum_{i \in \mathcal{N}} c_i x_{it} + c s_t + d z_t \right), \tag{21}$$

subject to

$$
\begin{aligned}
&\sum_{t=1}^{\tilde{\tau}(u)} s_t \geq 1, \\
&\sum_{t=\ell}^{\tau+\ell-1} s_t \geq 1, \qquad \ell = \max\{2, \tilde{\tau}(u) - \tau + 1\}, \ldots, T - \tau, \\
&\sum_{t=\ell}^{\tau+\ell-1} s_t \geq s_{\ell-1}, \qquad \begin{array}{c} \ell = 2, \ldots, \tilde{\tau}(u) - \tau, \\ \text{if } \ \tilde{\tau}(u) - \tau \geq 2, \end{array} \\
&s_t \leq z_t, \qquad t = 1, \ldots, T-1, \\
&\mathbf{x}_1 = \overline{\mathbf{x}}_1, \\
&s_t \in \mathbb{B}, \qquad t = 1, \ldots, T-1, \\
&(x, z) \in S,
\end{aligned}
$$

with S defined in (4). The computation of the second stage function (21) requires a discretization of the distribution G. Let $n \in \{1, \ldots, T-1\}$ and define $\kappa_n = \{k_1, \ldots, k_n\} \subseteq \{1, \ldots, T-1\}$ such that $k_{j+1} > k_j$, $j \in \{1, \ldots, n-1\}$. An n-discretization G_n of the distribution G has the probability mass function

$$
g_n(u) = \begin{cases} p_j, & \text{if } u = k_j, \quad j = 1, \ldots, n, \\ 0, & \text{if } u \notin \kappa_n. \end{cases} \tag{22}
$$

4.2.1 Error measure

The n-discretization (22) yields the replacement strategy

$$
\hat{\mathbf{x}}_1^n \in \underset{\mathbf{x}_1 \in \mathbb{B}^{N+2}}{\operatorname{argmin}} \, F_n(\mathbf{x}_1), \tag{23}
$$

with F_n defined as F in (20) with G replaced by G_n, so that $F_n(\mathbf{x}_1) = \sum_{j=1}^{n} f(\mathbf{x}_1, k_j) \cdot p_j$.

The distribution using the maximum number of support points in the model is G_{T-1}. Introduce the error measure for the expected cost between the two discretizations

$$
e(G_n, G_{T-1}) = F_{T-1}(\hat{\mathbf{x}}_1^n) - F_{T-1}(\hat{\mathbf{x}}_1^{T-1}), \tag{24}
$$

for $n \in \{1, \ldots, T-1\}$. The upper bound $e(G_n, G_{T-1}) \leq 2 \cdot \sup_{\mathbf{x}_1 \in \mathbb{B}^{N+2}} |F_n(\mathbf{x}_1) - F_{T-1}(\mathbf{x}_1)| \leq C \cdot \sup_{u \in \mathbb{R}} |G_n(u) - G_{T-1}(u)|$, for some $C \geq 0$, was derived in Svensson [19].

4.2.2 Discretization approaches

When discretizing $G(u), u \in [0, \infty)$ using $n \leq T-1$ points of support, the following questions arise: 1) How many points of support should we use? 2) Which points of support $\kappa_n \subseteq \kappa_{T-1}$ should we choose? 3) How should we distribute 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 [19] four discretization approaches were described: (i) Minimizing the sup-norm distance, (ii) using means in brackets, (iii) minimizing the Wasserstein distance, and (iv) moment preserving discretization; we refer to Svensson [19] for more details and present only the numerical results here.

4.2.3 Test results

The life of the stochastic component is modeled as

$$
\breve{G}(u) = 1 - e^{-(\frac{1}{\theta}u)^{\alpha}}, \quad u \geq 0, \tag{25}
$$

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

We model an engine with one stochastic and one deterministic component. In each time step there are 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 (23) with $n \in \{1, \ldots, 10\}$ points of support.

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 means method (with Approximation 1, cf. Svensson [19]).

The most accurate discretization possible has $T - 1 = 30$ points of support, one in every time point, for which an optimal replacement strategy was calculated according to (23) with $n = T - 1$. The difference between the two discretizations, using the error measure (24), was calculated. The values for the parameters and remaining lives of the components used are $\tilde{\tau}(u) = 4$ and 6, $\tau = 6$ and 10, $c = 60, 70, 100, 130$ and 150, $d = 70, 100$ and 150 and $\alpha = 1$ and 2.

The results in Figure 8 indicate that minimizing the Wasserstein and the sup norm distance seem preferable to using the moment preserving method. The main conclusions seem to be that using two points of support is worse than using one (the expected value) of the distribution, furthermore that there is a large gain in using three points of support and that there is not a large gain in using 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. A possible explanation for this is that the variance of the Weibull distribution (25) decreases with an increasing value of the parameter α, for constant θ, and for a distribution with large variance many points of support are needed for a good approximation of the distribution.

Finally, in Svensson [19] it was shown that many details were replaced even when not needed, which can to be due to the approximation made in that (narrow scenario tree) scheme: 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.

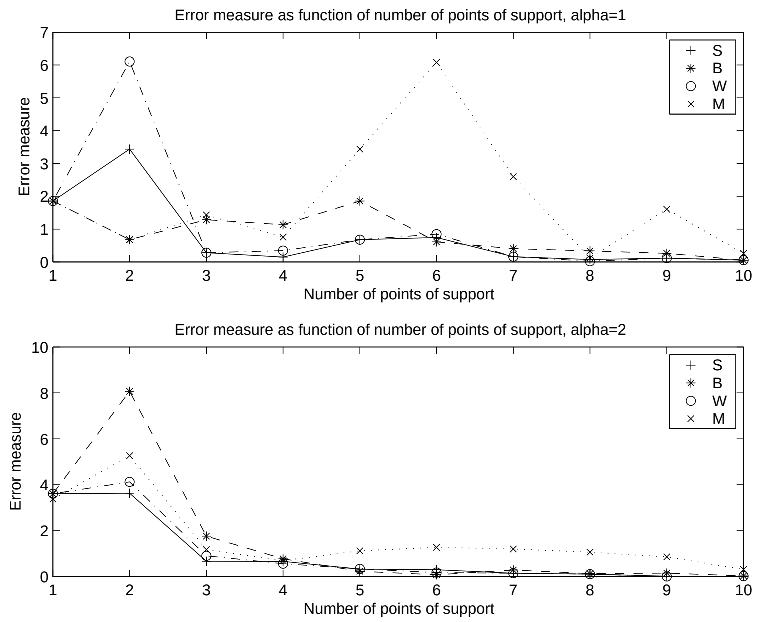

Figure 8: *The vertical axis represents the mean error measure with parameter values given above. The horizontal axis represents the number of support points. S, B, W and M refer, respectively, to the method minimizing the sup norm distance, the bracket method, the method minimizing the Wasserstein distance, and the moment preserving method.*

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 failure times, with respect to the problem of finding models that accurately describe the process of failure times. The main objective 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 types of stochastic processes (the non-homogeneous Poisson process and the non-stationary renewal process) to see which was the better fitting.

The second 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 developed 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 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 from only one 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, $t_1, \ldots, t_k$ say, as independent random variables the resulting multivariate discretized distribution will have $n_1 \cdot \ldots \cdot n_k$ 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 (cf. Section 1.3). In Section 6 we evaluate these policies against the optimization model through stochastic simulations.

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 $c_i < d$, then the policy dictates that it 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 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.

An age replacement policy is defined as follows. Each component $i \in \mathcal{N}$ is 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 $\delta = 0, 1, \ldots, T$ 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. Among its ten parts, four are safety critical (SC) (exhaust frame, roller, conical shaft, and air seal), while the remaining six (stator, seal and nozzle segments, case, disk, and blade) are on condition (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. The four methods then fare as follows. In relation to the total cost of using no opportunistic maintenance (below referred to as "None"), the other methods yield a cost of 66% [integer model (3), or "Integer"], 72% (age policy, or "Age"), and respectively 146% (value policy, or "Value). The total number of maintenance occasions are 14 ("None"), 4 ("Integer"), 3 ("Age"), and 9 ("Value"); each of the opportunistic methodologies reduce the former number considerably.

Figure 9 shows for each component how many individuals are replaced for each of the four methods.

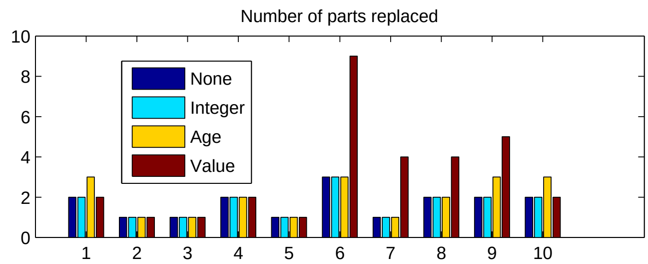

Figure 9: *Number of components replaced for the deterministic problem.*

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.) The value policy reduces the number of maintenance occasions 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 similarly apply the method of using no opportunistic maintenance.

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 10 summarizes the experiments where we have, for components 1, 4, 5, 6, 9, and 10 let a common value of β vary among the values 6, 4, and 2, thus gradually increasing the level of uncertainty.

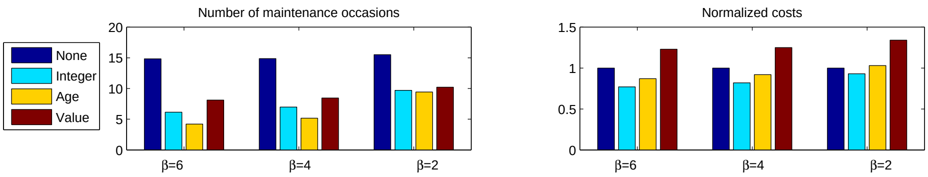

Figure 10: *Number of maintenance occasions and cost when the β value is the same across parts.*

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 the components have different β values, which is the most likely scenario; let components 1 and 4 have $\beta = 2$, components 5 and 6 have $\beta = 4$, and components 9 and 10 have $\beta = 6$. In relation to the total cost of using no opportunistic maintenance ("None"), the other methods then yield a cost of 83% ["Integer"], 96% (age policy, or "Age"), and respectively 126% ("Value). The total number of maintenance occasions are 15.01 ("None"), 7.73 ("Integer"), 6.07 ("Age"), and 8.35 ("Value").

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

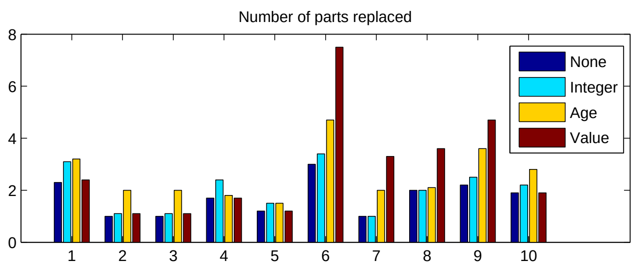

Figure 11: *Number of components replaced when the β value varies among parts.*

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 is however only marginally better.

In summary, 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, 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; the effectiveness of the age replacement policy is however problem dependent—it is not difficult to construct examples when this heuristic provides schedules that are 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 settings 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 treat this problem.

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, Department of Mathematics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 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] R. Barlow and F. Proschan, *Mathematical Theory of Reliability*, John Wiley & Sons, New York, NY, 1965.

[5] M. Berg and B. Epstein, *A modified block replacement policy*, Naval Research Logistics Quarterly, 23 (1976), pp. 15–24.

[6] 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.

[7] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver, *Combinatorial Optimization*, John Wiley & Sons, New York, NY, 1998.

[8] D. R. Cox, *Regression models and life-tables*, Journal of the Royal Statistical Society. Series B. Methodological, 34 (1972), pp. 187–220. With discussion by F. Downton, R. Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, S. Howard, B. Benjamin, J. J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, J. Kalbfleisch, R. L. Prentice, and N. Breslow, and a reply by D. R. Cox.

[9] 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 (1997), pp. 411–435.

[10] 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.

[11] D. B. Edmunds, *Modular engine maintenance concept considerations for aircraft turbine engines*, Aircraft Engineering, 50 (1978), pp. 14–17.

[12] 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.

[13] ——, *An optimal structured policy for maintenance of partially observable aircraft engine components*, Naval Research Logistics, 45 (1998), pp. 335–352.

[14] L. Mann, A. Saxena, and G. M. Knapp, *Statistical-based or condition-based preventive maintenance?*, Journal of Quality in Maintenance Engineering, 1 (1995), pp. 46–59.

[15] G. L. Nemhauser and L. A. Wolsey, *Integer and combinatorial optimization*, John Wiley & Sons, New York, NY, 1988.

[16] Y. Pawitan, *In All Likelihood*, Clarendon press, Oxford, UK, 2001.

[17] J. Svensson, *Two statistical models used on aircraft engine data modeling times between repairs*, technical report, Department of Mathematics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2005.

[18] ——, *Assessment of residual life based on retarding crack growth*, technical report, Department of Mathematics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2006.

[19] ——, *Discrete approximations of life distributions in optimal replacement*, technical report, Department of Mathematics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2007.

[20] R. Webster, *Convexity*, Oxford University Press, Oxford, UK, 1994.