Reliability optimization for series systems under uncertain component failure rates in the design phase

Citation for published version (APA):

Ge, Q., Peng, H., van Houtum, G. J. J. A. N., & Adan, I. J. B. F. (2018). Reliability optimization for series systems under uncertain component failure rates in the design phase. *International Journal of Production Economics*, 196(2), 163-175. <https://doi.org/10.1016/j.ijpe.2017.10.022>

DOI:

[10.1016/j.ijpe.2017.10.022](https://doi.org/10.1016/j.ijpe.2017.10.022)

Document status and date:

Published: 01/02/2018

Document Version:

Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Reliability Optimization for Series Systems under Uncertain Component Failure Rates in the Design Phase

Qianru Ge^{a,*}, Hao Peng^c, Geert-Jan van Houtum^{a,*}, Ivo Adan^{a,b}

^a*Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands*

^b*Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

^c*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*

Abstract

We develop an optimization model to determine the reliability design of critical components in a serial system. The system is under a service contract, and a penalty cost has to be paid by the OEM when the total system down time exceeds a predetermined level, which complicates the evaluation of the expected cost under a given reliability design. Furthermore, in the design phase for each critical component, all possible designs are subject to uncertain component failure rates. Considering the computational intractability of evaluating the system performance, we develop approximate evaluation methods which take the system uncertainty into account. Numerical results show that the method which includes randomness in the number of failures, failure rates and repair times leads to efficient and accurate evaluations and to close-to-optimal design decisions when used in an enumeration procedure for the optimization problem. We also show that ignoring these three types of uncertainty may result in bad design decisions.

Key words: Capital goods, Reliability optimization, Performance-based contracting, Life cycle cost

1. Introduction

Capital goods are machines or products that are used by manufacturers to produce their end-products or that are used by service organizations to deliver their services. Advanced technical systems such as medical systems, manufacturing systems, and defense systems, are examples of capital goods that are critical for the operational processes of their customers. System downtime of these capital goods can have serious consequences (e.g., millions of euros of reduced production output, extra waiting time of passengers,

*Corresponding author: T: +31 40 247 2230

Email addresses: q.ge@tue.nl (Qianru Ge), haopng@gmail.com (Hao Peng), G.J.v.Houtum@tue.nl (Geert-Jan van Houtum), i.adan@tue.nl (Ivo Adan)

failure of military missions) and maintaining these high-tech systems is too challenging for customers to take care of by themselves. Original equipment manufactures (OEMs) can take care of the maintenance and guarantee high system availability levels. These OEMs can be seen as performance providers rather than only solution providers (Helander and Möller, 2007). The guaranteed system availability levels are generally specified in Service Level Agreements (SLAs) within service contracts. Customers pay a price for the service contracts. When an OEM fails to meet the predetermined level of availability, the OEM needs to pay a penalty cost to the customer. Different types of service contracts are mentioned in Cohen et al. (2006), among which performance-based contracts (PBCs). According to Guajardo et al. (2012), “Performance-based contracting compensates the supplier based on the same outcome that the customer cares about (i.e., product utilization), and hence the supplier is motivated to increase product performance, associated with metrics such as product reliability and availability”. Therefore PBCs are a certain type of service contracts, in which system performance translates into financial bonuses and penalties (Selviaridis and Wynstra, 2015). In the main model studied in this paper, we include the penalty side of a PBC contract. We also show how the bonus side can be included (see Section 6.1).

Under a SLA on system availability, one of the OEM’s major concerns is the life cycle cost (LCC), defined as the total cost incurred in the design/development, production, operation, maintenance, support, and final disposition of a system over its anticipated useful life span (Barringer and Weber, 1996). There is literature aiming at cost saving under such a SLA (Öner et al., 2010). Measurements reported in Öner et al. (2007) for an engineer-to-order system show that the sum of maintenance cost and downtime cost is larger than the acquisition cost and constitutes a significant portion of the LCC. The service cost is incurred by system failures which are highly determined by system designs. Therefore, it is important that the LCC is taken into account in the reliability design decision.

Customers of capital goods measure the availability of these complex systems at the end of service contract periods. In the literature (Al Hanbali and Van der Heijden, 2013, De Souza E Silva and Gail, 1986 and Takacs, 1957), this availability during a service contract period is denoted as the *interval availability*. The realized interval availability of a capital good should meet the required performance levels. When the interval availability is below its target, or equivalently, when the interval downtime is above the target downtime, the OEMs will pay a penalty cost to the customer that is generally proportional to the exceedance. Therefore, it is important for the OEMs to have a method to determine the exceedance of the target downtime under a given design. This method can then be used when optimizing the reliability design.

In reality, engineers have to select a certain design from all possible alternatives for each critical component in a system during the design phase. The outcome of any development process for a certain design is uncertain with respect to its reliability level. For example, since the failure mechanisms of some emerging technologies (e.g., micro-electro-mechanical systems) are complex, it is often difficult to predict the actual reliability levels of the critical components before the development. Therefore newly-designed devices have been found to have component failure rates that deviate significantly from the expectations after the completion of a design. Brown and Burke (2000) and Brown et al. (2004) showed how individual system failure rates can deviate from the average failure rate of a population of similar equipment by collecting empirical data from the power delivery industry. This *uncertainty in component failure rates* also needs to be considered in the reliability design.

In this paper, we study the system reliability design problem by minimizing the LCC and considering both uncertain component reliability and interval availability under a SLA on system availability. So far, this problem has not been studied in the literature. For this new problem, our contribution is as follows. The key issue is the calculation of the exceedance of the target downtime under a given reliability design. In this calculation, one needs to take three uncertainties into account: the uncertainty in the component failure rates, the stochastic nature of the number of failures during a service contract period, and the stochastic repair times. This may be done by simulation, but that leads to a relatively long computation time per evaluation of a reliability design and to too long computation times for the optimization problem for instances of a reasonable size. Therefore, we propose a fast approximate evaluation method that takes all three uncertainties into account. This method is called the *full-uncertainty method*. We compare its performance to the performance of two other methods: a stochastic method ignoring the uncertain failure rates (partial method), and a deterministic method using expected total downtime as the actual total downtime (zero-uncertainty method). These approximate evaluation methods are also used in an optimization procedure for the reliability design, and hence lead to heuristic solutions for the reliability design problem. We show that the full-uncertainty method clearly outperforms the other two methods, both for the evaluation of a given design and when used in an enumeration procedure for the optimization problem. The use of the other two methods for the optimization problem may lead to bad reliability design decisions.

The remainder of the paper is organized as follows. In Section 2, we briefly review related literature. Section 3 gives the model description and model formulation. We describe the three approximate evaluation methods in Section 4. The numerical results are presented in Section 5. In Section 6, we first discuss how the bonus side of a PBC

contract can be included. Next, we show the extension to a contract period with multiple subperiods and an interval availability target per subperiod. Conclusions and directions for future research are presented in Section 7.

2. Literature

We review the literature regarding the main characteristics of our problem: Interval availability, reliability optimization, and uncertain component failure rates.

As defined by Nakagawa and Goel (1973), *interval availability* is the fraction of time that a system is operational during a time period $[0, T]$. In the literature, several methods have been proposed to measure interval availability during a service period. Takacs (1957) considers a general stochastic on-off process and derives an exact expression for the distribution function of the total length of the off periods during a time period $[0, T]$. These off periods may represent the availability of a technical system. The exact expression has a high computational complexity, so that it can only be used for smaller problem instances. In Van der Heijden (1988), an approximate evaluation of the interval availability distribution is developed based on two-moment fits for the on and off periods. This method is accurate and can solve large problem instances in short computation times. De Souza E Silva and Gail (1986) consider a Markov process on a finite state space, where a subset of states represents 'good states' and they derive an exact expression for the distribution function of the total amount of time that the Markov process is in good states during a time period $[0, T]$. The good states may represent states in which a given technical system works. Also this exact expression of De Souza E Silva and Gail (1986) has a high computational complexity, so that it only can be used for smaller problem instances. Al Hanbali and Van der Heijden (2013) consider a two-echelon spare parts inventory system consisting of a single depot and multiple bases, and they study the logistics availability of technical systems installed at the bases during a time period $[0, T]$. They derive an approximate Markov process where a subset of states corresponds with logistics availability. They follow a numerical approach that can solve large problem instances within reasonable computation times. For more references and approaches on interval availability, we refer to the literature discussion in Al Hanbali and Van der Heijden (2013).

The evaluation of a given reliability design is related to the interval availability literature. Like Takacs (1957) and De Souza E Silva and Gail (1986), we derive an exact formula for exceedance of the target downtime, which is equivalent to deriving the distribution function of the interval availability. Our formula is different because we assume that the off periods are very short and they occur according to Poisson processes. In addition, we have to take uncertainty in the failure rates into account. Next we develop approximate,

numerical approaches in order to be able to evaluate large problem instances. In these numerical approaches, we use two-moment fits, which is a general technique that was also used by Van der Heijden (1988). Nevertheless, our approach differs because it is based on a different exact formula.

Figure 1: Positioning of this paper

Regarding *reliability optimization*, a lot of work has been done in this area since the 1990s (Kuo and Wan (2007)). For example, Mettas and Kallenberg (2000) determined the minimum required reliability for each component of a system in order to achieve a system reliability goal with minimum cost. The cost function for each component in this paper has been used in other papers as well; see e.g. Huang et al. (2007), Öner et al. (2010), and Jin and Wang (2012). Many papers maximize the system reliability by different techniques. For example, a random search process has been proposed by Beraha and Misra (1974) to determine the optimal reliability for each stage of a multi-stage system. Hwang (1975) used sequential unconstrained minimization, and Li and Haimmes (1992) developed a 3-level decomposition approach to allocate the resources among subsystems optimally.

Some papers also built reliability allocation models to find optimal warranty policies for systems sold with traditional warranty contracts. For example, to minimize the system LCC, Monga and Zuo (1998) used genetic algorithms to solve the optimization problem and Öner et al. (2010) introduced a decision support model to jointly optimize the reliability level and spare parts inventory level of a single-component system. To max-

imize the profit, Huang et al. (2007) proposed a model to compute the optimal warranty policy under different market situations by using the maximum principle method. Park et al. (2015) formulated a model to determine the optimal warranty period for the manufacturer. They minimize the warranty cost consisting of the repair cost and downtime cost. In recent years a growing number of papers built reliability optimization models within the context of service contract. Jin and Tian (2012) developed a model to optimize the reliability design and inventory level. They minimize the life cycle cost under a non-stationary demand rate, and consider a dynamic stocking policy. Selçuk and Ağralı (2013) jointly optimized reliability levels of components and stock levels of spare parts for a multi-item system. They minimize the sum of holding and emergency shipment cost subject to different types of service constraints. Jin et al. (2015) integrated spare part inventory, maintenance and repair capacity into one model. They maximize the utility of both the supplier and the customer within a game-theoretic framework. They use a gradient-based heuristic and a hybrid algorithm to solve the problem for a single-item system and multi-item system, respectively. Öner et al. (2013) minimize the LCC of a number of identical systems by jointly optimizing the component redundancy and spare parts inventory.

In most previous work on the system reliability optimization problem (see e.g. Kuo and Wan, 2007, Jin and Wang, 2012, Kim et al., 2017), all parameters are assumed to be precise. However in reality there is considerable uncertainty and inaccuracy in the estimation of model parameters. In particular, one has *uncertain reliability*. To measure the uncertain system reliability due to insufficient component-level failure data, Coit (1997), Jin and Coit (2001), and Ramirez-Marquez and Levitin (2008) estimated the confidence intervals of the system reliability for different systems. Hussain and Murthy (1998) developed a model to determine the optimal redundancy design considering quality uncertainty due to manufacture variability, and obtained the best tradeoff between the manufacturing cost and warranty cost. In Hussain and Murthy (2003), the uncertain component failure rate has been modeled by a reliability growth model, and an optimization model has been proposed to find the best trade-off between the development cost and the warranty cost. Feizollahi and Modarres (2012) proposed a reliability design framework to address the uncertain component failure rates by using the Min-Max regret (also known as robust deviation) approach and transformed the nonlinear programming formulation to a linear binary version to get the exact solutions. None of the above works considered interval availability.

In this paper, we focus on the cases in which all three aspects in Figure 1 are considered. We propose a reliability optimization model with uncertain component failure rates and approximately evaluate the LCC under an interval availability constraint (as

specified in a service contract).

3. Model

Notation

$I, I $	set and number of critical component in the system
$J_i, J_i $	set and number of alternative designs of the i -th component
T	length of the service period
x_{ij}	decision variable for whether design j of component i is selected or not
$\mathbf{x}_i =$	$(x_{i1}, x_{i2}, \dots, x_{i J_i })$, the alternative that has been chosen for component i
$\mathbf{x} =$	$(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{ I })$, the selection of a design for each component
$B_{ij}, B_i(\mathbf{x}_i)$	random repair time for design j of component i and for component i
$B_i^{(s)}(\mathbf{x}_i)$	Sum of s independent repair times for component i
r_{ij}, v_{ij}	mean and standard deviation of B_{ij}
$S_{ij}, S_i(\mathbf{x}_i)$	number of failures in $[0, T]$ of design j of component i and of component i
$\Lambda_{ij}, \Lambda_i(\mathbf{x}_i)$	failure rate of design j of component i and of component i
$f_{\Lambda_{ij}}(\cdot), \mu_{ij}, \sigma_{ij}$	probability density function, mean and standard deviation of Λ_{ij}
$D_{ij}, D_i(\mathbf{x}_i), D(\mathbf{x})$	total downtime of design j of component i , of component i and of the whole system in $[0, T]$
D_0	target downtime of the system
c_{ij}^r	repair cost for each failure of design j of component i
c^p	penalty rate for the exceeded downtime
$c_{ij}^a, A_i(\mathbf{x}_i), A(\mathbf{x})$	acquisition cost of design j of component i , of component i and of the system
$R_i(\mathbf{x}_i), R(\mathbf{x})$	expected repair cost of component i and of the system in $[0, T]$
$P(\mathbf{x})$	expected system penalty cost in $[0, T]$
$\pi(\mathbf{x})$	expected LCC in $[0, T]$

During the design phase of a single system, engineers have to select a certain design from all possible alternatives for each critical component in the system. Suppose the system is comprised of a set of critical components I ; the components are numbered $1, 2, \dots, |I|$. If one of these critical components fails, the system as a whole stops working. For each critical component, one design needs to be selected from possible alternatives denoted by J_i ; these alternatives are numbered $1, 2, \dots, |J_i|$. Each design candidate in J_i has its own uncertain failure rate and its own cost and repair parameters. We aim to find the optimal combination of designs for the system so that the expected LCC over the service period $[0, T]$ of a service contract is minimized (so, T denotes the length of the service period).

We assume that the failure process of each design for each component is independent of other components and their designs, and it follows a Poisson process. This implies that the lifetimes of the components are independent and exponentially distributed. For a certain design j of component i ($i \in I, j \in J_i$), we denote its failure rate as Λ_{ij} . The outcome of any development process for a certain design is uncertain. Therefore, the failure rate Λ_{ij} of design j for component i is usually not known for sure before the development of the

design. We use $f_{\Lambda_{ij}}(\cdot)$ as the probability density function of the random failure rate Λ_{ij} before the development of the design, which reflects the prior belief/information about the failure rate uncertainty of the technologies used in the design. The expectation and standard deviation of Λ_{ij} are denoted by μ_{ij} and σ_{ij} ; it is assumed that $\mu_{ij} > 0$. In the evaluation of the expected LCC over the service period $[0, T]$, these design uncertainties will be taken into account.

The system will be sold together with a SLA over a service period $[0, T]$. The OEM is responsible for all repairs in $[0, T]$. Moreover the total downtime of the service period should be lower than a predetermined level D_0 . A penalty cost will be paid by the OEM to compensate the customer if the total downtime exceeds D_0 . As a result, the expected total cost of a system over $[0, T]$ consists of three parts: (a) acquisition cost, (b) repair cost, and (c) penalty cost.

Define the binary decision variable x_{ij} as

$$x_{ij} = \begin{cases} 1 & \text{if design } j \text{ for component } i \text{ is selected,} \\ 0 & \text{otherwise.} \end{cases}$$

The OEM selects one design from all possible candidates for each component, so we have $\sum_{j \in J_i} x_{ij} = 1$. The vector of x_{ij} -s of a given component i , $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{i|J_i|})$ denotes which component design has been chosen for component i . The failure rate of the chosen design for component $i \in I$ is denoted by $\Lambda_i(\mathbf{x}_i)$. The vector of $\mathbf{x}_i$ -s, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|I|})$, denotes the selection of a design for each component. The expected LCC, denoted by $\pi(\mathbf{x})$, is the sum of the total design cost $A(\mathbf{x})$, the expected system repair cost $R(\mathbf{x})$ and the expected penalty cost $P(\mathbf{x})$. These functions are further developed in Subsections 3.1-3.3.

Due to the randomness of the failure rates and the occurrence of failures according to Poisson processes, the total life cycle cost is random. We assume that the OEM is risk-neutral (Kim et al., 2007) and hence aims to minimize the expected LCC $\pi(\mathbf{x})$. Our optimization problem is formulated as:

$$\begin{aligned} \text{(P)} \quad & \min_{\mathbf{x}} \quad \pi(\mathbf{x}) \\ \text{s.t.} \quad & \sum_{j \in J_i} x_{ij} = 1, \quad \text{for all } i \in I \\ & x_{ij} \in \{0, 1\}, \quad \text{for all } i \in I, j \in J_i \end{aligned}$$

where $\pi(\mathbf{x}) = A(\mathbf{x}) + R(\mathbf{x}) + P(\mathbf{x})$.

Remark 1. *The service period, referring to the contract period of the OEM and its*

customer, is assumed to cover a large part of the life time of the system. Its length is T , and may be 20-30 years. The service period itself is denoted by the interval $[0, T]$. LCC is defined as the sum of all costs incurred in $[0, T]$. We assume that the acquisition cost is incurred at time 0. After the service period, the system may still be used for another period, but this can depend on various factors such as the state of the system, how well the system meets the requirement of the customer at that time, and the alternative systems that can be bought at that time. We assume that the value of the system at the end of the service period is negligible relative to all costs included in our model.

Remark 2. We assume constant failure rates. For electronic parts, as mentioned in Minges (1989), this assumption is reasonable. Hence, our model fits for systems with mainly electronic parts. Our model may also fit for systems with components that degrade over time, such as most mechanical parts. If the increase in the failure rates is limited for all components and no preventive maintenance is applied, then assuming constant failure rates may still be appropriate. For each component, the failure rate can be taken equal to the average failure rate during the lifetime of the component. If some components would have a strongly increasing failure rate, then it is likely that preventive maintenance is applied and then our model would not fit (i.e., one would have to develop a more advanced model that models the preventive maintenance policy and in which the preventive maintenance policy is adapted over time depending on the realizations of lifetimes of the component).

3.1. Acquisition cost

Let c_{ij}^a denote the cost of designing and manufacturing component i according to design j ($i \in I, j \in J_i$). For the design, it includes all involved cost elements, e.g., for human resources, experimental equipment, and testing or prototype units. The acquisition cost for component i is given by

$$A_i(\mathbf{x}_i) = \sum_{j \in J_i} c_{ij}^a x_{ij},$$

and the total system acquisition cost is given by

$$A(\mathbf{x}) = \sum_{i \in I} A_i(\mathbf{x}_i) = \sum_{i \in I} \sum_{j \in J_i} c_{ij}^a x_{ij}. \quad (1)$$

Obviously, this formula enables a simple exact evaluation of the total acquisition cost under a given design.

3.2. Repair cost

When a failure occurs in period $[0, T]$, a repair will be performed by the OEM. Let c_{ij}^r be the repair cost for each failure of the j -th design for component i ($i \in \{1, 2, \dots, |I|\}$,

$j \in \{1, 2, \dots, |J_i|\}$. The repair cost c_{ij}^r corresponds to diagnosis cost, replacement cost, and other service costs for each repair. Let $S_i(\mathbf{x}_i)$ denote the total number of failures for component i during $[0, T]$. The expected number of repairs for component i during $[0, T]$ is given by:

$$\mathbb{E}[S_i(\mathbf{x}_i)] = \mathbb{E}[\Lambda_i(\mathbf{x}_i)T] = \sum_{j \in J_i} \mathbb{E}(\Lambda_{ij})Tx_{ij} = \sum_{j \in J_i} \mu_{ij}Tx_{ij},$$

where $\Lambda_i(\mathbf{x}_i) = \sum_{j \in J_i} \Lambda_{ij}x_{ij}$ is the random failure rate of component i given a certain design implied by vector $\mathbf{x}_i$. The expected repair cost for component i is expressed by:

$$R_i(\mathbf{x}_i) = \sum_{j \in J_i} \mathbb{E}(\Lambda_{ij})Tc_{ij}^r x_{ij} = \sum_{j \in J_i} \mu_{ij}Tc_{ij}^r x_{ij}.$$

The expected system repair cost in $[0, T]$ is given by:

$$R(\mathbf{x}) = \sum_{i \in I} R_i(\mathbf{x}_i) = \sum_{i \in I} \sum_{j \in J_i} \mu_{ij}Tc_{ij}^r x_{ij}. \quad (2)$$

Obviously, this formula enables a simple exact evaluation of the total repair cost under a given design.

3.3. Penalty cost

We assume that the repair time per failure of component i with design j ($i \in I$, $j \in J_i$), B_{ij} , is a generally distributed random variable with mean r_{ij} and standard deviation v_{ij} . We assume that this random repair time is negligibly small compared to the service period $[0, T]$ (the B_{ij} are in the order of hours while the service period $[0, T]$ is in the order of years). Hence the repair time can be assumed to have no effect on the total number of failures of a component during service period $[0, T]$. Implicitly, we already made use of this property when we determined the repair cost.

When the total system downtime over the service period $[0, T]$ exceeds D_0 , a penalty cost has to be paid by the OEM to customers with a rate c^p (D_0 is at the same time scale as B_{ij}). We allow that the system downtime of each failure varies among different components with different designs. Hence the total system downtime $D(\mathbf{x})$ over the service period $[0, T]$ depends on the number of failures $S_i(\mathbf{x}_i)$ and the repair time per failure $B_i(\mathbf{x}_i)$ for component $i \in I$ in $[0, T]$. Notice that the number of failures of component i , $S_i(\mathbf{x}_i)$, is a Poisson distributed random variable. Hence, the distribution of $S_i(\mathbf{x}_i)$ is given as

$$\begin{aligned} \mathbb{P}[S_i(\mathbf{x}_i) = s_i] &= \sum_{j \in J_i} x_{ij} \mathbb{P}(S_{ij} = s_i) \\ &= \sum_{j \in J_i} x_{ij} \int_0^\infty \mathbb{P}(S_{ij} = s_i | \Lambda_{ij} = \lambda) f_{\Lambda_{ij}}(\lambda) d\lambda \\ &= \sum_{j \in J_i} x_{ij} \int_0^\infty \frac{e^{-\lambda T} (\lambda T)^{s_i}}{s_i!} f_{\Lambda_{ij}}(\lambda) d\lambda, \end{aligned}$$

where S_{ij} is the number of failures of component $i \in I$ with design $j \in J_i$. The expected penalty cost due to exceeding the target downtime D_0 is given as

$$P(\mathbf{x}) = \mathbb{E} \left\{ \left[D(\mathbf{x}) - D_0 \right]^+ c^p \right\} = \sum_{s_1=0}^{\infty} \sum_{s_i=0}^{\infty} \cdots \sum_{s_{|I|=0}^{\infty} \prod_{i \in I} \mathbb{P} \left[S_i(\mathbf{x}_i) = s_i \right] \mathbb{E} \left[\sum_{i \in I} B_i^{(s_i)}(\mathbf{x}_i) - D_0 \right]^+ c^p. \quad (3)$$

This formula for $P(\mathbf{x})$ contains an $|I|$ -fold summation and multiple integrations over the ranges of the random failure rates are needed per term of the summation (to obtain the probabilities $\mathbb{P}[S_i(\mathbf{x}_i) = s_i]$). This becomes computationally intractable even for a small number of components, for $|I| \geq 5$, say. Therefore, in the next section, approximate evaluation methods are developed.

4. Approximate evaluation

In this section, we are going to explain how to approximately evaluate a given design for the whole system. For a given $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|I|})$, the design cost $A(\mathbf{x})$ and the expected repair cost $R(\mathbf{x})$ can be determined by (1) and (2). These two cost terms are linear functions of the decision variables $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|I|})$.

For the exact evaluation of the expected penalty cost $P(\mathbf{x})$, we will suffer from the “curse of dimensionality” when the number of critical components is larger than 5 say, since each critical component contributes a dimension in computing the summation and its terms in (3). From (3), we can also observe that there are three types of uncertainty existing in $D(\mathbf{x})$ throughout $[0, T]$. The first type of uncertainty originates from the number of failures per component under a given design with a given failure rate, which is a Poisson distributed random variable. The second level of uncertainty comes from the failure rates Λ_{ij} ($i \in I, j \in J_i$), which are generally distributed random variables. The third type of uncertainty is due to the stochastic repair times B_{ij} ($i \in I, j \in J_i$). We now introduce three approximation methods to estimate the expected penalty cost, which vary in dealing with the uncertainty types of $D(\mathbf{x})$. In the first evaluation method, called *zero-uncertainty method*, we use the expected downtime as the actual downtime, thus ignoring all three types of uncertainty. In the current literature, the uncertainty existing in the failure rates is often ignored. So, in the second evaluation method, called *partial-uncertainty method*, we only consider the uncertainty of the number of failures of the system and the stochastic repair times of each component and we assume the expected failure rate for the selected design per component. In the third evaluation method, called *full-uncertainty method*, we take all three uncertainty types of $D(\mathbf{x})$ into account.

All three evaluation methods are based on expressions for the first two moments of $D(\mathbf{x})$. They are obtained as follows. For the total downtime in $[0, T]$ of the j -th design

for component i , denoted by D_{ij} , the first moment is given by:

$$\mathbb{E}(D_{ij}) = \int_0^\infty \int_0^\infty \mathbb{E}[D_{ij} | \Lambda_{ij} = \lambda, B_{ij} = b] f_{\Lambda_{ij}}(\lambda) f_{B_{ij}}(b) d\lambda db = \int_0^\infty \lambda T b f_{\Lambda_{ij}}(\lambda) f_{B_{ij}}(b) d\lambda db = \mu_{ij} T r_{ij} \quad (4)$$

For the second moment and variance of D_{ij} , we obtain (we use the property that, under a given failure rate, D_{ij} is the sum of all compounds of a compound Poisson process in an interval $[0, T)$):

$$\begin{aligned} \mathbb{E}(D_{ij}^2) &= \int_0^\infty \mathbb{E}[D_{ij}^2 | \Lambda_{ij} = \lambda, B_{ij} = b] f_{\Lambda_{ij}}(\lambda) f_{B_{ij}}(b) d\lambda db \\ &= \int_0^\infty \left\{ \lambda T \mathbb{E}(B_{ij})^2 + \lambda^2 T^2 [\mathbb{E}(B_{ij})]^2 \right\} f_{\Lambda_{ij}}(\lambda) d\lambda \\ &= r_{ij}^2 (\mu_{ij}^2 T^2 + \sigma_{ij}^2 T^2 + \mu_{ij} T) + v_{ij}^2 \mu_{ij} T \\ \text{Var}(D_{ij}) &= \mathbb{E}(D_{ij})^2 - [\mathbb{E}(D_{ij})]^2 = r_{ij}^2 (\sigma_{ij}^2 T^2 + \mu_{ij} T) + v_{ij}^2 \mu_{ij} T \end{aligned} \quad (5)$$

Then, for component i , we have:

$$\begin{aligned} D_i(\mathbf{x}_i) &= \sum_{j \in J_i} D_{ij} x_{ij}, \\ \mathbb{E}[D_i(\mathbf{x}_i)] &= \sum_{j \in J_i} r_{ij} \mu_{ij} T x_{ij}, \\ \text{Var}[D_i(\mathbf{x}_i)] &= \sum_{j \in J_i} \left[r_{ij}^2 (\sigma_{ij}^2 T^2 + \mu_{ij} T) + v_{ij}^2 \mu_{ij} T \right] x_{ij}. \end{aligned} \quad (6)$$

Finally, for the total downtime of the system in $[0, T]$, we obtain:

$$\begin{aligned} D(\mathbf{x}) &= \sum_{i \in I} D_i(\mathbf{x}_i) = \sum_{i \in I} \sum_{j \in J_i} D_{ij} x_{ij}, \\ \mathbb{E}[D(\mathbf{x})] &= \sum_{i \in I} \mathbb{E}[D_i(\mathbf{x}_i)] = \sum_{i \in I} \sum_{j \in J_i} r_{ij} \mu_{ij} T x_{ij}, \end{aligned} \quad (7)$$

$$\text{Var}[D(\mathbf{x})] = \sum_{i \in I} \text{Var}[D_i(\mathbf{x}_i)] = \sum_{i \in I} \sum_{j \in J_i} \left[r_{ij}^2 (\sigma_{ij}^2 T^2 + \mu_{ij} T) + v_{ij}^2 \mu_{ij} T \right] x_{ij}. \quad (8)$$

4.1. Zero-uncertainty method

Under the zero-uncertainty method, the downtime $D(\mathbf{x})$ is simply approximated by its expected value:

$$D_{\text{zero}}(\mathbf{x}) = \sum_{i \in I} \sum_{j \in J_i} r_{ij} \mu_{ij} T x_{ij}.$$

Then the expected excess downtime $D^E(\mathbf{x})$ and penalty cost $P(\mathbf{x})$ are approximated as:

$$\begin{aligned} D_{\text{zero}}^E(\mathbf{x}) &= \left[D_{\text{zero}}(\mathbf{x}) - D_0 \right]^+, \\ P_{\text{zero}}(\mathbf{x}) &= \left[D_{\text{zero}}(\mathbf{x}) - D_0 \right]^+ c^p. \end{aligned}$$

4.2. Partial-uncertainty method

Under the partial-uncertainty method, the failure rates Λ_{ij} ($i \in I$, $j \in |J_i|$) are approximated by their means μ_{ij} . This is equivalent to pretending that all σ_{ij} are equal to 0. The expressions for $\text{Var}(D_{ij})$, $\text{Var}[D_i(\mathbf{x}_i)]$, and $\text{Var}[D(\mathbf{x})]$ then simplify to (see (5), (6), (8)):

$$\begin{aligned}\text{Var}(D_{ij}) &= (v_{ij}^2 + r_{ij}^2)\mu_{ij}T, \\ \text{Var}[D_i(\mathbf{x}_i)] &= \sum_{j \in J_i} (v_{ij}^2 + r_{ij}^2)T\mu_{ij}x_{ij}, \\ \text{Var}[D(\mathbf{x})] &= \sum_{i \in I} \sum_{j \in J_i} (v_{ij}^2 + r_{ij}^2)T\mu_{ij}x_{ij}\end{aligned}\tag{9}$$

By (7) and (9), the coefficient of variation of $D(\mathbf{x})$ is given as (notice that both $\mathbb{E}[D(\mathbf{x})]$ and $\text{Var}[D(\mathbf{x})]$ are strictly positive): $c_v(D(\mathbf{x})) = \frac{\sqrt{\text{Var}[D(\mathbf{x})]}}{\mathbb{E}[D(\mathbf{x})]}$.

We now fit a mixed Erlang distribution on the mean $\mu_D = \mathbb{E}[D(\mathbf{x})]$ and coefficient of variation $c_v^D = c_v(D(\mathbf{x}))$, and then we compute the expected excess downtime based on the fitted distribution. For the two-moment fit, we use an Erlang($k-1, k$) distribution if $0 < c_v^D \leq 1$ and a hyperexponential distribution otherwise. This is further explained below.

Suppose $0 < c_v^D \leq 1$. Then we fit an Erlang($k-1, k$) with parameters with parameters (k, θ, q) , $k \in \mathbb{N}$, $k \geq 2$, $\theta > 0$, $0 < q \leq 1$, on μ_D and c_v^D . We get such a fit via the following formulas (see Tijms (1994), pp. 335-337). The k is chosen such that $1/k < (c_v^D)^2 \leq 1/(k-1)$. Next, q and θ are taken equal to

$$q = \frac{1}{1 + (c_v^D)^2} \left[k(c_v^D)^2 - \sqrt{k(1 + (c_v^D)^2) - k^2(c_v^D)^2} \right], \quad \theta = \frac{k - q}{\mu_D}$$

Then the expected excess downtime can be approximated as (see Appendix B.2):

$$D_{part}^E(\mathbf{x}) = \left(\frac{k - q}{\theta} - D_0 \right) \sum_{j=0}^{k-2} \frac{(\theta D_0)^j}{j!} e^{-\theta D_0} + \left(\frac{k - q}{\theta} \right) \frac{(\theta D_0)^{k-1}}{(k-1)!} e^{-\theta D_0}.$$

Suppose $c_v^D \geq 1$. Then we fit an hyperexponential distribution with parameters (θ_1, θ_2, q) , $\theta_1 > 0$, $\theta_2 > 0$, $0 < q \leq 1$, on μ_D and c_v^D . We get such a fit via the following formulas (see Tijms (1994), pp. 335-337):

$$\theta_1 = \frac{2}{\mu_D} \left(1 + \sqrt{\frac{(c_v^D)^2 - \frac{1}{2}}{(c_v^D)^2 + 1}} \right), \quad \theta_2 = \frac{4}{\mu_D} - \theta_1, \quad q = \frac{\theta_1(\theta_2 \mu_D - 1)}{\theta_2 - \theta_1}.\tag{10}$$

Then the expected excess downtime can be approximated as (see Appendix B.2):

$$D_{part}^E(\mathbf{x}) = \frac{q}{\theta_1} e^{-\theta_1 D_0} + \frac{1 - q}{\theta_2} e^{-\theta_2 D_0}$$

Finally, $P(\mathbf{x})$ is approximated by: $P_{part}(\mathbf{x}) = D_{part}^E(\mathbf{x})c^p$.

4.3. Full-uncertainty method

Under the full-uncertainty method, we take both the uncertainty of the failure rates as well as the number of failures into account. We use the formulas (7) and (8) for the mean and variance of $D(\mathbf{x})$. Next, we determine the corresponding coefficient of variation by $c_v(\mathbf{x}) = \frac{\sqrt{\text{Var}[D(\mathbf{x})]}}{\mathbb{E}[D(\mathbf{x})]}$.

We now fit a mixed Erlang distribution on the mean $\mu_D = \mathbb{E}[D(\mathbf{x})]$ and coefficient of variation $c_v^D = c_v(D(\mathbf{x}))$, and then we compute the expected excess downtime based on the fitted distribution. These steps are the same as under the partial-uncertainty method. The resulting approximation for the expected excess downtime is denoted by $D_{full}^E(\mathbf{x})$. Finally, $P(\mathbf{x})$ is approximated by: $P_{full}(\mathbf{x}) = D_{full}^E(\mathbf{x})c^p$.

5. Computational results

In this section, we execute two numerical experiments to test the accuracy of our approximate evaluation methods. In Section 5.1, we investigate the quality of the approximate evaluation results generated by the zero-uncertainty method, partial-uncertainty method, and full-uncertainty method under given designs; here, we compare the approximate results to simulated results. In Section 5.2, we investigate the quality of the heuristic policies generated when solving Problem (P) via enumeration and approximate evaluations.

5.1. Experiment 1

5.1.1. Test bed

To assess the quality of the three approximate evaluation methods under given designs, we compare results of these methods with simulated results. We use a Monte Carlo simulation method; see Appendix A for more details on this method. The essence of the approximate evaluation methods is that they approximate the expected excess downtime $\mathbb{E}\{(D(\mathbf{x}) - D_0)^+\}$. For the rest, no approximations are made. Hence, we compare the evaluation methods with simulation on the basis of the generated values for $\mathbb{E}\{(D(\mathbf{x}) - D_0)^+\}$.

We use a full factorial design for the test bed. We vary three factors: the number of the critical components $|I|$, the coefficient of variation c_v for the component failure rates and the so-called downtime budget factor D_f . Per instance, we have one design per component and thus one design for the whole system. As a result, in this experiment the notation can be simplified. We distinguish one stochastic failure rate Λ_i with mean μ_i and standard deviation σ_i per component $i \in I$. We assume that each Λ_i is lognormally distributed and choose the parameters of the lognormal distribution such that the mean and standard deviation are equal to μ_i and σ_i respectively. Further, we have one mean repair time r_i and one corresponding standard deviation v_i per component $i \in I$. In

Table 1: Parameter settings for $|I|$, c_v , and D_f in Experiment 1

$ I $	c_v	D_f
5, 25, 50, 75, 100	0.2, 0.5, 0.8, 1.1, 1.4	1, 1.05, 1.1, 1.15, 1.2, 1.25, 1.3

each instance of the test bed, we take values for μ_i and r_i that vary over the components $i \in I$. For the standard deviations of the failure rates Λ_i , we use c_v ; we take $\sigma_i = c_v \mu_i$ for all $i \in I$. We assume that the standard deviations v_i for the repair times are equal to 0, i.e., that the repair times are deterministic. In Section 5.1.3, we will execute an additional experiment with stochastic repair times. The factor D_f is used to generate different values of D_0 by the following expression:

$$D_0 = D_f \sum_{i \in I} \mu_i r_i T, \quad (11)$$

where $\sum_{i \in I} \mu_i r_i T$ is the expected total downtime. Thus, if D_f is one, the target downtime D_0 is set equal to the expected total downtime. The length of the service period is chosen equal to $T = 10$ years.

The test bed consists of 175 instances that are obtained by taking all combinations of the values for $|I|$, c_v , and D_f as shown in Table 1. The values for μ_i (in failures per year) and r_i (in hours) are given in Table 2. For the zero-uncertainty method, it is

Table 2: Parameter settings for μ_i and r_i , $i \in I$, in Experiment 1

μ_i for $i = 1, \dots, I $	r_i for $i = 1, \dots, I $
$\{0.200, \dots, 0.200 - (i-1)\frac{0.18}{n-1}, \dots, 0.02\}$	$\{1, 3, 5, 1, 3, 5, \dots\}$

apparent that when $D_f \geq 1$, $D_{zero}^E = 0$. Hence, this is true for all instances. For the partial-uncertainty method, D_{part}^E does not depend on the values for the σ_i , and thus we get the same outcome for all instances that only differ with respect to the value for c_v .

5.1.2. Results

Per instance of the test bed, we determine the approximations D_{zero}^E , D_{part}^E , and D_{full}^E via the three evaluation methods and we determine D_{sim}^E via simulation. The 95% confidence intervals of all the simulation results are relatively small (the width of a confidence interval is at most 0.001 times the simulated value D_{sim}^E , and hence we can use D_{sim}^E to determine the accuracy of the approximations). Each of the approximations and the simulated excess downtime is normalized by D_0 , which gives:

$$\hat{D}_k^E = \frac{D_k^E}{D_0}, \quad k \in \{zero, part, full, sim\}.$$

Per instance, the accuracy of each approximation is assessed by the absolute gaps for the normalized excess downtime:

$$Gap = |\hat{D}_k^E - \hat{D}_{sim}^E|, \quad k \in \{zero, part, full\}.$$

The results for these Gaps are presented in Table 3. In this table, the *Avg Gap* and *max Gap* columns show the average gap and maximum gap of each method for a group of instances that has the same value of one of the three factors. For example, the values 0.93% and 3.20% in the first row are the average and maximum gaps of the full-uncertainty method in the 35 instances with $|I| = 5$ components in the system. In the last row, the results for the whole set of 175 instances are given.

For the zero-uncertainty method, the average gap is 6.70% and the max gap is 31.93% for all instances. The average and max gap for the partial-uncertainty method is 2.39% and 12.17%, respectively. For the full-uncertainty method, the average gap is 0.25% and the maximum gap is 3.20%. The results show that the full-uncertainty method gives accurate results in general, while the zero-uncertainty method leads to a poor approximation on average. The partial-uncertainty method performs significantly worse than the full-uncertainty method. The computation time per instance of the simulation method is ranging from less than 2 minutes to 76.18 hours with an average number of 2.30 hours. For the approximation methods, the computation times are less than 0.03 seconds for each of the instances. All numerical results are run on a computer with Intel Core 2 Duo E7500 processor at 2.93GHz and 4 gigabytes physical RAM available (plus 4 gigabytes virtual memory).

Table 3: Average gap and maximum gap for the three approximate evaluation methods in Experiment 1

		zero-uncertainty method		partial-uncertainty method		full-uncertainty method	
		<i>Avg Gap</i> (%)	<i>max Gap</i> (%)	<i>Avg Gap</i> (%)	<i>max Gap</i> (%)	<i>Avg Gap</i> (%)	<i>max Gap</i> (%)
$ I $	5	18.32	31.93	5.44	12.17	0.93	3.20
	25	6.41	15.49	2.58	6.80	0.17	0.87
	50	3.81	11.19	1.68	5.06	0.07	0.43
	75	2.76	9.22	1.25	4.21	0.05	0.28
	100	2.19	8.04	1.00	3.70	0.04	0.20
D_f	1	11.91	31.93	3.12	12.17	0.34	3.20
	1.05	9.36	28.63	2.95	11.79	0.30	3.06
	1.1	7.43	25.78	2.70	11.42	0.26	2.87
	1.15	5.96	23.26	2.41	11.02	0.23	2.69
	1.2	4.85	21.03	2.10	10.59	0.22	2.52
	1.25	4.01	19.07	1.84	10.17	0.21	2.32
	1.3	3.36	17.31	1.61	9.73	0.20	2.16
	1.4	2.76	15.49	1.25	8.41	0.17	1.68
c_v	0.2	4.44	20.62	0.14	0.87	0.07	0.39
	0.5	5.22	22.73	0.92	2.98	0.05	0.29
	0.8	6.47	25.79	2.16	6.03	0.09	0.39
	1.1	7.93	28.95	3.62	9.20	0.33	1.57
	1.4	9.43	31.93	5.12	12.17	0.72	3.20
All		6.70	31.93	2.39	12.17	0.25	3.20

Under all three methods, the gaps become smaller for increasing values of $|I|$. This is due to the fact that the total downtime is the sum of $|I|$ independent random variables. The coefficient of variation of the total downtime decreases for increasing $|I|$, and thus

the effect of the uncertainties of the failure rates Λ_i and the resulting numbers of failure in the service period $[0, T]$ decreases. For $|I| \rightarrow \infty$, the total downtime converges to the expected downtime $\sum_{i \in I} \mu_i r_i T$ in distribution; this also holds for the approximations, and thus all gaps become 0 in that case. When considering the gaps for increasing values of D_f , we see that they decrease under all three methods, but the decrease is much stronger for the zero-uncertainty method than for the other two methods. When considering the gaps for increasing values of c_v , we see that they increase under all three methods, but now the increase is much less for the full-uncertainty method than for the other two methods.

Figure 2: The probability density function for the total downtime as obtained via simulation and the full-uncertainty method for $|I| = 50$ and $c_v = 1.1$.

Figure 3: The probability density function for the total downtime as obtained via simulation and the full-uncertainty method for $|I| = 5$ and $c_v = 1.4$.

In order to better demonstrate the accuracy of the full-uncertainty method, we also compare the probability density functions and cumulative probability functions of the total downtime as obtained by simulation with the ones estimated by the full-uncertainty method; see Figure 2 and Figure 3. As we can see in Figure 2, generally, the full-uncertainty method gives an accurate approximation for the simulated curve for the probability density functions as well as the cumulative probability functions of the total

downtime for various settings of $|I|$ and c_v . Only when $|I| = 5$ and $c_v = 1.4$, the shape of the approximation curve for the probability density function is different from the shape of the related simulation curve when the total downtime is smaller than 20 hours. However, the approximation for the cumulative probability function is reasonable for all values of the total downtime.

5.1.3. Stochastic vs. deterministic repair time

In this section we investigate the effect of stochastic instead of deterministic repair times on the results of the previous section.

Table 4: Parameter settings for $|I|$, c_v , and D_f in Experiment 1'

$ I $	c_v	D_f
5, 25, 50	0.2, 0.8, 1.4	1, 1.1, 1.2, 1.3

We use a subset of the test bed described in Section 5.1.1, see Table 4. There are 36 instances in the test bed. To simulate all instances, for stochastic repair times, we assume that the repair time per component B_i is exponentially distributed with rate r_i , $i \in I$ (i.e., $v_i = r_i$ for all $i \in I$). The results for the stochastic and deterministic repair time test bed are shown in Table 5 and Table 6, respectively.

Table 5: Average gap and maximum gap for the three approximate evaluation methods with stochastic repair time in Experiment 1'

		zero-uncertainty method		partial-uncertainty method		full-uncertainty method	
		<i>Avg Gap</i>	<i>max Gap</i>	<i>Avg Gap</i>	<i>max Gap</i>	<i>Avg Gap</i>	<i>max Gap</i>
		(%)	(%)	(%)	(%)	(%)	(%)
$ I $	5	24.37	39.75	5.52	12.38	0.77	2.07
	25	9.90	20.01	3.26	7.76	0.95	1.72
	50	6.40	14.78	2.41	6.12	0.87	1.77
D_f	1	20.22	39.75	4.29	12.38	0.87	1.68
	1.1	14.76	32.73	3.97	11.30	0.87	1.77
	1.2	10.96	27.20	3.56	10.32	0.86	1.92
	1.3	8.28	22.81	3.11	9.45	0.85	2.07
c_v	0.2	9.66	26.65	0.51	1.83	0.66	2.07
	0.8	13.24	32.82	3.08	5.45	0.69	1.06
	1.4	17.76	39.75	7.60	12.38	1.24	1.77
All		13.55	39.75	3.73	12.38	0.86	2.07

The results in Tables 5 and 6 show that the gaps under stochastic repair times are somewhat higher than under deterministic times. But, for the rest, we get the same conclusions as in Section 5.1.2. The full-uncertainty method produces accurate approximations under all parameter settings, while the zero-uncertainty method has the largest average and maximum gap. The full-uncertainty method significantly outperforms the

partial-uncertainty method. Furthermore, the gaps of the zero-uncertainty and partial-uncertainty method and the average gap of the full-uncertainty method are changing in the exact same directions. That is, the gap decreases for increasing values of $|I|$ and D_f and increases for increasing values of c_v .

Table 6: Average gap and maximum gap for the three approximate evaluation methods with deterministic repair in Experiment 1'

		zero-uncertainty method		partial-uncertainty method		full-uncertainty method	
		<i>Avg Gap</i>	<i>max Gap</i>	<i>Avg Gap</i>	<i>max Gap</i>	<i>Avg Gap</i>	<i>max Gap</i>
		(%)	(%)	(%)	(%)	(%)	(%)
$ I $	5	18.58	31.93	5.55	12.17	1.08	3.20
	25	6.69	15.49	2.67	6.80	0.21	0.87
	50	4.09	11.19	1.75	5.06	0.08	0.43
D_f	1	15.59	31.93	4.06	12.17	0.60	3.20
	1.1	10.67	25.78	3.64	11.42	0.49	2.87
	1.2	7.48	21.03	3.06	10.59	0.40	2.52
	1.3	5.41	17.31	2.52	9.73	0.35	2.16
c_v	0.2	6.67	20.62	0.21	0.87	0.11	0.39
	0.8	9.44	25.79	2.97	6.03	0.14	0.39
	1.4	13.24	31.93	6.78	12.17	1.12	3.20
All		9.79	31.93	3.32	12.17	0.46	3.20

The average computation time of the simulations for the stochastic repair time instances is 6.14 times more than for the deterministic repair time instances. That was a main reason to limit the analysis in this section to a subset of the whole test bed of Section 5.1.1. Because stochastic and deterministic repair times lead to the same conclusions and insights, we can limit ourselves to deterministic repair times in the rest of Section 5.

5.2. Experiment 2

In the second experiment, we solve Problem (P) by enumeration. In this enumeration, simulation or one of the approximate evaluation methods can be used to evaluate each given solution. The solution that is obtained via simulation is generally optimal or very close to optimal. The corresponding costs are used as estimate for the optimal costs. The use of each of the approximate evaluation methods leads to a heuristic solution for Problem (P). We investigate the quality of these heuristic solutions. In Section 5.2.1, we describe the test bed of Experiment 2. The quality of the heuristic solutions is discussed in Section 5.2.2. Finally, managerial insights are given in Section 5.2.3.

5.2.1. Test bed

For the test bed for Experiment 2, we use a full factorial design, in which 5 factors are varied. For all instances, we assume that $|I| = 10$, which is the largest $|I|$ that is

computationally tractable under the use of simulation. For the same reason, we have $|J_i| = 2$ designs per component $i \in I$. The five factors that are varied are as follows: the downtime factor D_f , the acquisition cost c_{ij}^a ($i \in I, j \in J_i$), the expected failure rate μ_{ij} ($i \in I, j \in J_i$), the coefficient of variation c_v , and the penalty rate c^p . The values of these five factors are given in Table 7. For each factor, we choose three values, so in total 243 instances are obtained. The target downtime D_0 (in hours) is set equal to the product of D_f and the average value of the expected downtimes where the same weight is taken for all designs, i.e., $D_0 = D_f \sum_{i \in I} \frac{1}{|J_i|} \sum_{j \in J_i} \mu_{ij} r_{ij} T$, with $T = 10$. All repair times are assumed to be deterministic and their means r_{ij} are taken equal to $r_{ij} = 3$ hours ($i \in I, j \in J_i$). The length of the service period is fixed at $T = 10$ years. The standard deviations σ_{ij} are set at $\sigma_{ij} = c_v \mu_{ij}$ ($i \in I, j \in J_i$). The cost per repair c_{ij}^r ($i \in I, j \in J_i$) is equal to 30% of the corresponding acquisition cost c_{ij}^a .

Table 7: Parameter settings for D_f , c_{ij}^a , μ_{ij} , c_v , and c_p in Experiment 2

Name of parameter	No. val.	Values
Downtime budget factor (D_f)	3	1, 1.2, 1.4
Acquisition cost (c_{ij}^a)	3	$\left\{ \begin{array}{l} (1) \begin{cases} c_{i1}^a: [500, \dots, 500i, \dots, 5000] \\ c_{i2}^a: 1.5 \times [500, \dots, 500i, \dots, 5000] \end{cases} \\ (2) \begin{cases} c_{i1}^a: [500, \dots, 500i, \dots, 5000] \\ c_{i2}^a: 2 \times [500, \dots, 500i, \dots, 5000] \end{cases} \\ (3) \begin{cases} c_{i1}^a: [500, \dots, 500i, \dots, 5000] \\ c_{i2}^a: 2.5 \times [500, \dots, 500i, \dots, 5000] \end{cases} \end{array} \right.$
Expected failure rates (μ_{ij})	3	$\left\{ \begin{array}{l} (1) \begin{cases} \mu_{i1}: [0.15, 0.14, 0.12, 0.08, 0.06, 0.16, 0.18, 0.2, 0.04, 0.02] \\ \mu_{i2}: [0.1, 0.12, 0.11, 0.06, 0.03, 0.13, 0.14, 0.15, 0.03, 0.01] \end{cases} \\ (2) \begin{cases} \mu_{i1}: [0.15, 0.14, 0.12, 0.08, 0.06, 0.16, 0.18, 0.2, 0.04, 0.02] \\ \mu_{i2}: 0.75 \times [0.1, 0.12, 0.11, 0.08, 0.03, 0.13, 0.14, 0.15, 0.03, 0.01] \end{cases} \\ (3) \begin{cases} \mu_{i1}: [0.15, 0.14, 0.12, 0.08, 0.06, 0.16, 0.18, 0.2, 0.04, 0.02] \\ \mu_{i2}: 0.5 \times [0.1, 0.12, 0.11, 0.08, 0.03, 0.13, 0.14, 0.15, 0.03, 0.01] \end{cases} \end{array} \right.$
Coefficient of variation (c_v)	3	0.3, 0.9, 1.5
Penalty cost factor (c^p)	3	5000, 10000, 20000

In our test bed, the acquisition cost of the second design of every component is higher than for the first design. So, we refer to the second design as the "expensive" design and to the first design as the "cheap" design. Notice that for the expensive design of every component, the average and standard deviation of the failure rate Λ_{ij} are always smaller than the ones for the cheap design. As a result, there is an incentive in all cases to choose for the expensive design.

5.2.2. Results

For each instance, we solve Problem (P) by enumeration, where 4 different methods are used for the evaluations of given solutions: simulation and each of the three approximate evaluation methods. The solution that is obtained under the use of simulation is denoted by $\mathbf{x}_{sim}^*$. This solution will generally be optimal or close to optimal. We use

the corresponding cost $\pi(\mathbf{x}_{sim}^*)$ as an estimate for the optimal cost. Under the use of the zero-, partial-, and full-uncertainty method, we obtain a heuristic solution $\mathbf{x}_{zero}^*$, $\mathbf{x}_{part}^*$, and $\mathbf{x}_{full}^*$, respectively. The corresponding total cost equals $\pi(\mathbf{x}_{zero}^*)$, $\pi(\mathbf{x}_{part}^*)$, and $\pi(\mathbf{x}_{full}^*)$, where this total cost is evaluated by simulation. For each of the heuristic solutions we determine the relative optimality gap

$$OG_k = \frac{\pi(\mathbf{x}_k^*) - \pi(\mathbf{x}_{sim}^*)}{\pi(\mathbf{x}_{sim}^*)}, \quad k \in \{zero, part, full\}.$$

The results for the optimality gaps are given in Table 8. Per line, we give the averages over the instances with the same value for one of the 5 factors that is varied. At the last line, we give the averages over all instances. Per heuristic, we also give the percentage of instances in which the heuristic solution $\mathbf{x}_k^*$, $k \in \{zero, part, full\}$, deviates from $\mathbf{x}_{sim}^*$ (for at least one component). This percentage is denoted by NOS_k , where NOS stands for "No Optimal Solution". Further, we consider the percentage of components for which the expensive design is chosen under solution $\mathbf{x}_k^*$, $k \in \{zero, part, full\}$, which we denote by $PExp_k$. We compare this percentage with the corresponding percentage $PExp_{sim}$ under $\mathbf{x}_{sim}^*$. The absolute difference is denoted by

$$\Delta PExp_k = |PExp_k - PExp_{sim}|, \quad k \in \{zero, part, full\}.$$

The averages of these absolute differences are also listed in Table 8. The computation times per instance under the use of simulation are in the order of 60-334 hours, while they are all less than one second under the use of the approximate evaluation methods.

Table 8: Results on the quality of the heuristic solutions $\mathbf{x}_{zero}^*$, $\mathbf{x}_{part}^*$, and $\mathbf{x}_{full}^*$

		Zero-uncertainty			Partial-uncertainty			Full-uncertainty		
		OG_{zero} (%)	NOS_{zero} (%)	$\Delta PExp_{zero}$ (%)	OG_{part} (%)	NOS_{part} (%)	$\Delta PExp_{part}$ (%)	OG_{full} (%)	NOS_{zero} (%)	$\Delta PExp_{full}$ (%)
D_f	1	105.7	100	71	0.31	34.6	4	0.02	13.6	1
	1.2	83.9	100	61	1.32	54.3	8	0.02	12.3	1
	1.4	60.1	100	49	3.21	64.2	11	0.03	14.8	2
c_{ij}^a	(1)	110.3	100	73	1.09	45.7	7	0.03	17.3	2
	(2)	79.2	100	59	1.63	53.1	8	0.02	9.9	1
	(3)	60.2	100	48	2.12	54.3	8	0.02	13.6	1
μ_{ij}	(1)	21.9	100	47	1.21	49.4	9	0.06	28.4	3
	(2)	72.5	100	65	1.66	46.9	8	0.02	11.1	1
	(3)	155.3	100	68	1.97	56.8	7	0.00	1.2	0
c^p	5000	31.8	100	45	0.96	50.6	7	0.03	17.3	2
	10000	72.4	100	61	1.81	56.8	9	0.03	13.6	1
	20000	145.6	100	74	2.07	45.7	8	0.02	9.9	1
c_v	0.3	80.9	100	54	0.04	14.8	1	0.01	4.9	0
	0.9	84.9	100	61	1.42	67.9	9	0.01	14.8	1
	1.5	83.9	100	65	3.38	70.4	3	0.05	21.0	2
All		83.2	100	60	1.61	51.0	8	0.02	13.6	1

The results in Table 8 show that the average optimality gap of the zero-uncertainty method is large. Under this method, one never obtains the optimal solution (i.e., one gets always another solution than $\mathbf{x}_{sim}^*$) and chooses much less frequently for the expensive

designs of components than under $\mathbf{x}_{sim}^*$. The average optimality gap of the partial-uncertainty method is only 1.61%, while in 51.0% of the instances another solution is obtained than $\mathbf{x}_{sim}^*$. The average optimality gap of the full-uncertainty method is 0.02%, which is very low. This method gives the same solution as $\mathbf{x}_{sim}^*$ in 86.4% of the instances and only in 13.6% of the instances a slightly different solution is obtained. We may conclude that the full-uncertainty method generates the best heuristic solutions. This method performs very well, both in terms of the total cost of the generated solution and in terms of the generated solution itself. Hence, this method is very appropriate to solve Problem (P).

5.2.3. Managerial insights

In this section, we further analyze the close-to-optimal solutions $\mathbf{x}_{full}^*$ generated by the full-uncertainty method. Extra results of these solutions are given in Table 9 and Table 10. In Table 9, first of all we list $PExp_{full}$, the percentage of components for which the expensive design is chosen. We give the average, maximum, and minimum value over all (last line) or subgroups (other lines) of instances. Next, we give results for PPC_{full} , which denotes the probability that a penalty cost has to be paid, i.e., the probability that the total downtime exceeds the target D_0 . We also give results for the expected exceeded downtime $D_{full}^E(\mathbf{x}_{full}^*)$ and for the same downtime but then divided by the target value D_0 . In Table 10, we consider the changes in the objective function under $\mathbf{x}_{full}^*$, i.e., in

Table 9: Analysis of the solutions $\mathbf{x}_{full}^*$ generated by the full-uncertainty method - Part 1

		$PExp_{full}$ (%)			PPC_{full} (%)			$D_{full}^E(\mathbf{x}_{full}^*)$ (hrs)			$D_{full}^E(\mathbf{x}_{full}^*)/D_0$ (%)		
		Avg.	max	min	Avg.	max	min	Avg.	max	min	Avg.	max	min
D_f	1	70.37	90.00	70.37	25.99	55.62	25.99	2.62	8.60	2.62	9.21	28.26	9.21
	1.2	61.11	90.00	10.00	17.76	38.76	3.52	1.78	6.49	0.16	5.22	17.75	0.55
	1.4	50.49	80.00	10.00	12.71	29.21	1.98	1.28	4.91	0.09	3.24	12.92	0.27
c_{ij}^a	(1)	73.83	90.00	20.00	15.12	35.36	1.98	1.44	5.42	0.09	4.52	17.81	0.27
	(2)	59.14	80.00	10.00	18.82	49.79	3.07	1.88	7.54	0.14	5.83	24.76	0.43
	(3)	49.01	80.00	10.00	22.52	55.62	5.34	2.36	8.60	0.34	7.32	28.26	1.00
μ_{ij}	(1)	48.52	90.00	10.00	27.30	55.62	6.74	3.20	8.60	0.41	9.33	28.26	0.95
	(2)	65.68	90.00	20.00	18.57	44.77	3.29	1.73	4.91	0.17	5.59	14.52	0.44
	(3)	67.78	90.00	30.00	10.59	28.46	1.98	0.75	2.22	0.09	2.74	9.30	0.27
c^p	5000	45.80	80.00	10.00	23.47	55.62	3.07	2.49	8.60	0.14	7.74	28.26	0.43
	10000	61.85	80.00	10.00	17.95	43.80	3.07	1.76	6.28	0.14	5.45	20.63	0.43
	20000	74.32	90.00	20.00	15.04	34.79	1.98	1.43	5.25	0.09	4.48	17.26	0.27
c_v	0.3	53.83	90.00	10.00	17.92	55.62	1.98	1.27	5.47	0.09	3.99	17.96	0.27
	0.9	60.86	90.00	10.00	18.55	51.35	2.54	1.77	6.93	0.14	5.51	22.75	0.42
	1.5	67.28	90.00	10.00	20.00	46.05	5.08	2.64	8.60	0.40	8.17	28.26	1.21
All		60.66	90.00	10.00	18.82	55.62	1.98	1.89	8.60	0.09	5.89	28.26	0.27

$\pi(\mathbf{x}_{full}^*)$ in comparison to the optimal cost π_{ref}^* that has been obtained for a reference instance. The reference instance is defined as the instance where the second value is chosen for each of the parameters listed in Table 7. For this reference instance, the values

of π_{ref}^* , and the underlying cost terms A_{ref}^* , R_{ref}^* and P_{ref}^* are 64666, 43000, 9998 and 11669, respectively. We list values for the ratio of the total cost of a given instance and the total cost in the reference instance. This ratio is equal to $\pi(\mathbf{x}_{full}^*)/\pi_{ref}^*$. Similarly, we list ratios for the underlying cost components, i.e., we list values for $A(\mathbf{x}_{full}^*)/A_{ref}^*$, $R(\mathbf{x}_{full}^*)/R_{ref}^*$, and $P(\mathbf{x}_{full}^*)/P_{ref}^*$.

Table 10: Analysis of the solutions $\mathbf{x}_{full}^*$ generated by the full-uncertainty method - Part 2

		$\pi(\mathbf{x}_{full}^*)/\pi_{ref}^*$ (%)			$A(\mathbf{x}_{full}^*)/A_{ref}^*$ (%)			$R(\mathbf{x}_{full}^*)/R_{ref}^*$ (%)			$P(\mathbf{x}_{full}^*)/P_{ref}^*$ (%)		
		avg.	max	min	avg.	max	min	avg.	max	min	avg.	max	min
D_f	1	123	271	68	99	127	66	93	159	51	240	901	19
	1.2	103	216	65	92	127	65	90	152	54	154	619	17
	1.4	90	177	63	85	127	65	87	148	54	108	432	6
c_{ij}^a	(1)	92	234	63	84	90	66	76	100	51	138	879	6
	(2)	106	253	67	94	106	65	92	129	69	165	901	25
	(3)	117	271	67	99	127	66	102	159	84	198	901	55
μ_{ij}	(1)	124	271	66	84	127	65	105	159	88	289	901	49
	(2)	105	197	67	95	127	67	95	122	75	149	521	26
	(3)	87	137	63	97	127	74	70	88	51	63	223	6
c^p	5000	87	127	63	82	111	65	85	120	54	107	369	6
	10000	103	187	64	93	127	65	90	133	54	151	538	12
	20000	126	271	66	102	127	69	96	159	51	244	901	15
c_v	0.3	91	182	63	87	127	65	88	152	54	106	432	6
	0.9	103	219	65	92	127	65	90	152	51	154	633	14
	1.5	122	271	67	97	127	66	92	159	51	241	901	23
All		105	271	63	92	127	65	90	159	51	167	901	6

The results in Table 9 show that the expensive designs are chosen for 60.66% of the components, when we look at all instances of the test bed. The average probability PPC_{full} that a penalty cost has to be paid is 18.82%. The excess downtime normed by D_0 is on average 5.89%. The values of $PExp_{full}$ show that the expensive designs of components are more often chosen when: (i) D_f decreases (i.e., the target D_0 decreases); (ii) the differences $c_{i2}^a - c_{i1}^a$ in acquisition cost between the expensive and cheap design decrease (when going from (3) to (1) for the c_{ij}^a , see Table 7); (iii) the differences $\mu_{i1} - \mu_{i2}$ in expected failure rates increase (when going from (1) to (3) for the μ_{ij}); (iv) the penalty cost c^p increases; (v) the standard deviations σ_{ij} increase (i.e., the factor c_v increases). These effects can be explained as follows. When D_f decreases or c^p increases, there is a higher need for more reliable components. When the $c_{i2}^a - c_{i1}^a$ decrease or the $\mu_{i1} - \mu_{i2}$ increase, it becomes more attractive to choose the expensive designs. When the standard deviations σ_{ij} increase, the distribution of the total downtime $D(\mathbf{x})$ gets a higher coefficient of variation, and thus generally the penalty costs will increase under target downtimes D_0 that are higher than the expected downtime. We do have such target downtimes D_0 for our solutions $\mathbf{x}_{full}^*$; see the values for PPC_{full} in Table 9. Hence, increasing σ_{ij} leads to more excess downtime and thus a stronger need to choose for the expensive design for more components.

The effects of increasing/decreasing parameters on PPC_{full} are sometimes the same as for $PExp_{full}$ and sometimes they are opposite. In all cases, the effects on $D_{full}^E(\mathbf{x}_{full}^*)$ and on $D_{full}^E(\mathbf{x}_{full}^*)/D_0$ are the same as for PPC_{full} . That PPC_{full} , $D_{full}^E(\mathbf{x}_{full}^*)$, and $D_{full}^E(\mathbf{x}_{full}^*)/D_0$ decrease when c^p increases, does make sense. Under an increasing c^p more expensive designs are chosen and thus the downtime distribution decreases and so does the excess downtime and the probability that one has an excess downtime. Similarly, the effect of decreasing $c_{i2}^a - c_{i1}^a$ and increasing $\mu_{i1} - \mu_{i2}$ can be explained. The effects of a decreasing D_f and increasing c_v are not so easy to explain because of effects that work against each other (e.g., a decreasing D_f gives a decreasing downtime target D_0 and at the same time failure rates decrease because of more expensive designs that are chosen).

The results in Table 10 show that the costs $\pi(\mathbf{x}_{full}^*)$, which are close to the optimal costs, increase when: (i) D_f decreases (i.e., the target D_0 decreases); (ii) the acquisition cost factors c_{i2}^a increase (when going from (1) to (3) for the c_{ij}^a , see Table 7); (iii) the failure rates μ_{ij} increase (when going from (3) to (1) for the μ_{ij}); (iv) the penalty cost factor c^p increases; (v) the standard deviations σ_{ij} increase (i.e., the factor c_v increases). For the underlying cost terms, we see the same effects in general. Furthermore, when D_f decreases, we know that more expensive components are chosen, and thus the acquisition cost increases. We also see an increase in repair cost; the effect of the higher cost per repair appears to be stronger than the decrease in number of repairs. It also leads to a very strong increase in downtime cost, which is due to the stricter downtime requirement. When c_{i2}^a gets larger, although less expensive components are chosen, the acquisition cost still increases, as well as the penalty cost and repair cost. When μ_{i2} decreases, more expensive components are chosen, which results in increasing acquisition cost. Since the savings from the repair cost and penalty cost are larger than the increment of the acquisition cost, the total cost reduces with decreasing μ_{i2} . When c^p or c_v increases, more expensive components are chosen but the repair cost and penalty cost still increases.

6. Extensions

6.1. Including a bonus

As denoted in the introduction, in our model, we included the penalty side of a PBC contract. In this subsection, we discuss how our model can be adapted to include also the bonus side of a PBC contract. Let c^c be the compensation rate, where we may assume that $c^c \leq c^p$. The expected bonus under design $\mathbf{x}$ is denoted by $W(\mathbf{x}) = \mathbb{E}[D(\mathbf{x}) - D_0]^- c^c$. This function can be rewritten as

$$W(\mathbf{x}) = \mathbb{E}[D(\mathbf{x}) - D_0]^+ c^c - \mathbb{E}[D(\mathbf{x}) - D_0] c^c = P(\mathbf{x}) \frac{c^c}{c^p} - c^c \left(\sum_{i \in I} \sum_{j \in J_i} r_{ij} \mu_{ij} T x_{ij} - D_0 \right).$$

With the bonus being added, we get an optimization problem that is the same as Problem (P) but with a modified objective function $\pi'(\mathbf{x})$ instead of objective function $\pi(\mathbf{x})$. The modified objective function is equal to

$$\pi'(\mathbf{x}) = A(\mathbf{x}) + R(\mathbf{x}) + P(\mathbf{x}) - W(\mathbf{x}) = A(\mathbf{x}) + R(\mathbf{x}) + \frac{c^p - c_c}{c^p} P(\mathbf{x}) + c^c \left(\sum_{i \in I} \sum_{j \in J_i} r_{ij} \mu_{ij} T x_{ij} - D_0 \right).$$

This modified objective function is similar to the original objective function and hence the new optimization procedure can be solved in the same way as Problem (P).

6.2. Multiple subperiods with a target downtime per subperiod

In our model, we have assumed one long service period with a target downtime for that full period. In this subsection, we briefly discuss the effect of having multiple subperiods instead of one period, with a target downtime per subperiod.

Suppose that the contract period $[0, T]$ is divided into M ($\in \mathbb{N} = \{1, 2, \dots\}$) subperiods, which are numbered $1, \dots, M$. For subperiod m , the length is T_m (> 0) and the downtime target is D_0^m (> 0). The T_m and D_0^m are such that $\sum_{m=1}^M T_m = T$ and $\sum_{m=1}^M D_0^m = D_0$. The penalty costs for subperiod m , $P^m(\mathbf{x})$, are given by

$$P^m(\mathbf{x}) = \sum_{s_1=0}^{\infty} \cdots \sum_{s_i=0}^{\infty} \cdots \sum_{s_{|I|}=0}^{\infty} \prod_{i \in I} \mathbb{P} \left[S_i^m(\mathbf{x}_i) = s_i \right] \mathbb{E} \left[\sum_{i \in I} B_i^{(s_i)}(\mathbf{x}_i) - D_0^m \right]^+ c^p,$$

where $S_i^m(\mathbf{x}_i)$ is the number of failures of component i in subperiod m . The new LCC function is equal to

$$\pi''(\mathbf{x}) = A(\mathbf{x}) + R(\mathbf{x}) + \sum_{m=1}^M P_m(\mathbf{x}).$$

Figure 4: LCCs and the percentage of expensive components in the optimal solutions under different numbers of subperiods

This function replaces the objective function $\pi(\mathbf{x})$ in Problem (P). The penalty functions $P^m(\mathbf{x})$ can be approximated in the same way as the penalty function $P(\mathbf{x})$. The modified optimization problem can again be solved via enumeration.

To show the impact of having multiple subperiods instead of one large service period, we execute a small experiment. We use one instance of the full factorial test bed of Experiment 2. For each of the parameters specified in Table 7, we take the second option; i.e., we take $D_f = 1.2$, $c_v = 0.9$, $c^p = 10000$, and option (2) is taken for the c_{ij}^a and μ_{ij} . The number of

subperiods M is varied from 1 to 5, and we take subperiods of equal length and divide the downtime target equally over the subperiods, i.e., $T_m = T/M$ and $D_0^m = D_0/M$ for all $m = 1, \dots, M$. In the enumeration method, we use the full-uncertainty method for the evaluations. The results are given in Figure 4.

We see that the LCC increases strongly when the number of subperiods increases. The percentage of expensive components increases from 16.31% to 22.82% when going from 1 to 5 subperiods, which is also strong. The reason for both increases is as follows. Under more subperiods, the number of events (failures) per subperiod is much smaller and thus one gets less close to the expected downtime per subperiod. To avoid too high penalty costs, it helps to choose more expensive components which leads to less failures. But still the LCC increases a lot. This small experiment shows that an OEM has to be careful when guaranteeing a maximum downtime per subperiod instead of a maximum downtime for one large service period.

7. Conclusion

In this paper, we considered the reliability design for a serial system with uncertain component failure rates. The system is sold with a full service contract and hence it is relevant to minimize the total life cycle cost when deciding on the reliability design. The total life cycle cost includes the penalty cost for exceeding a given target downtime during the term of the service contract. In this problem, we distinguish three types of uncertainties, being the uncertain component failure rates themselves, the uncertainty in the number of failures during the service period under given failure rates, and the uncertain repair times. We developed the so-called full-uncertainty method which takes all three types of uncertainty into account and which is based on two-moment fits to determine the total downtime and the penalty cost. We have shown that this method: (i) gives fast and accurate approximations for the total cost under a given design; (ii) works well in combination with enumeration for the reliability design problem for which close-to-optimal solutions are generated. We also showed that this method outperforms two other methods.

In future research, we would like to extend our work to situations where the expected failure rate can be chosen from a given interval (i.e., where the alternative designs form a continuous set of designs).

Appendices

A. Procedures of the Monte Carlo simulation

Per instance of Experiment 1, the following simulation procedure is used:

Step 1 First, we generate the sequences of r_i , μ_i and σ_i , $i \in |I|$. Then we get D_0 immediately from (11). Then, for each component i , we take one sample $\hat{\Lambda}_i$ for $\Lambda_i \sim G(\mu_i, \sigma_i)$, where $G(\mu_i, \sigma_i)$ is a lognormal distribution with mean μ_i and standard deviation σ_i . Next, given that the number of failures of component i , denoted by S_i , is Poisson distributed with parameter $\hat{\Lambda}_i T$, we take one sample $\hat{S}_i$ for $S_i \sim \text{Pois}(\hat{\Lambda}_i T)$. This gives one simulation result $\hat{D}_{sim}^E$:

$$\hat{D}_{sim}^E = (\sum_{i \in I} \hat{S}_i r_i - D_0)^+.$$

Step 2 Step 1 is repeated 400000 times to get $\hat{D}_{sim,1}^E, \dots, \hat{D}_{sim,400000}^E$. Next, we determine the average $\bar{D}_{sim}^E = \sum_{k=1}^{400000} \hat{D}_{sim,k}^E$ of this series. This completes one subrun.

Step 3 Step 2 is repeated m times until a certain requirement* is met to get $\bar{D}_{sim,1}^E, \dots, \bar{D}_{sim,m}^E$. Next, we determine the average $D_{sim}^E = \frac{1}{m} \sum_{k=1}^m \bar{D}_{sim,k}^E$ of this series. The corresponding 95% percent confidence interval is given as:

$$(D_{sim}^E - t_{(m-1, 2.5\%)} \sqrt{\frac{S^2}{m}}, \quad D_{sim}^E + t_{(m-1, 2.5\%)} \sqrt{\frac{S^2}{m}}),$$

where $S^2 = (1/(m-1)) \sum_{k=1}^m (\bar{D}_{sim,k}^E - D_{sim}^E)^2$ and $t_{(m-1, 2.5\%)}$ is the upper 2.5% critical point for the t-distribution with $(m-1)$ degrees of freedom.

*: Step 2 is repeated until the width of the confidence interval divided by the average value is smaller than 0.001.

B. First partial moment for mixed Erlang distributions

B.1. Erlang($k-1, k$) distribution

Consider an Erlang($k-1, k$) distribution for a random variable X with parameters (k, θ, q) , $k \in \mathbb{N}$, $k \geq 2$, $\theta > 0$, $0 < q \leq 1$. The probability density function and the cumulative distribution function are given by

$$\begin{aligned} e_{k-1,k}(x) &= q\theta^{k-1} \frac{x^{k-2}}{(k-2)!} e^{-\theta x} + (1-q)\theta^k \frac{x^{k-1}}{(k-1)!} e^{-\theta x}, x \geq 0, \\ E_{k-1,k}(x) &= q \left(1 - \sum_{j=0}^{k-2} \frac{(\theta x)^j}{j!} e^{-\theta x} \right) + (1-q) \left(1 - \sum_{j=0}^{k-1} \frac{(\theta x)^j}{j!} e^{-\theta x} \right), x \geq 0. \end{aligned}$$

Let x_0 be a nonnegative constant. Then the first partial moment $\mathbb{E}[(X - x_0)^+]$ is equal to

$$\begin{aligned}\mathbb{E}[(X - x_0)^+] &= \int_0^\infty (x - x_0)^+ e_{k-1,k}(x) dx \\ &= \frac{q(k-1)}{\theta} \int_{x_0}^\infty \theta^k \frac{x^{k-1}}{(k-1)!} e^{-\theta x} dx + \frac{k(1-q)}{\theta} \int_{x_0}^\infty \theta^{k+1} \frac{x^k}{k!} e^{-\theta x} dx \\ &\quad - x_0 \int_{x_0}^\infty e_{k-1,k}(x) dx.\end{aligned}$$

Let the cumulative distribution function of a pure Erlang- k distribution with scale parameter $\theta > 0$ be denoted by $E_k^\theta(X) = 1 - \sum_{j=0}^{k-1} \frac{(\lambda x)^j}{j!} e^{-\lambda x}$, $x \geq 0$. Then the formula for $\mathbb{E}[(X - x_0)^+]$ can be further rewritten as

$$\begin{aligned}\mathbb{E}[(X - x_0)^+] &= \frac{q(k-1)}{\theta} \left[1 - E_k^\theta(x_0) \right] + \frac{k(1-q)}{\theta} \left[1 - E_{k+1}^\theta(x_0) \right] - x_0 \left[1 - E_{k-1,k}^\theta(x_0) \right] \\ &= \left(\frac{k-q}{\theta} - x_0 \right) \sum_{j=0}^{k-2} \frac{(\theta x_0)^j}{j!} e^{-\theta x_0} + \left(\frac{k-q}{\theta} \right) \frac{(\theta x_0)^{k-1}}{(k-1)!} e^{-\theta x_0}.\end{aligned}$$

B.2. Hyperexponential distribution

Consider a hyperexponential distribution for a random variable X with parameters (θ_1, θ_2, q) , $\theta_1 > 0$, $\theta_2 > 0$, $0 < q \leq 1$. The probability density function and the cumulative distribution function are given by

$$\begin{aligned}h_2(x) &= q\theta_1 e^{-\theta_1 x} + (1-q)\theta_2 e^{-\theta_2 x}, x \geq 0, \\ H_2(x) &= q(1 - e^{-\theta_1 x}) + (1-q)(1 - e^{-\theta_2 x}), x \geq 0.\end{aligned}$$

Let x_0 be a nonnegative constant. Then the first partial moment $\mathbb{E}[(X - x_0)^+]$ is equal to

$$\begin{aligned}\mathbb{E}[(X - x_0)^+] &= \int_0^\infty (x - x_0)^+ h_2(x) dx \\ &= q \int_{x_0}^\infty (x - x_0) \theta_1 e^{-\theta_1 x} dx + (1-q) \int_{x_0}^\infty (x - x_0) \theta_2 e^{-\theta_2 x} dx \\ &= \frac{q}{\theta_1} e^{-\theta_1 x_0} + \frac{1-q}{\theta_2} e^{-\theta_2 x_0}.\end{aligned}$$

8. References

References

- Al Hanbali, A., Van der Heijden, M., 2013. Interval availability analysis of a two-echelon, multi-item system. *European Journal of Operational Research* 228(3), 494-503.
- Barringer, H.P., Weber, D.P., 1996. Life cycle cost tutorial, in: Fifth International Conference of Process Plant Reliability, Gulf Publishing Company, Houston, USA.

- Beraha, D., Misra, K.B., 1974. Reliability optimization through random search algorithm. *Microelectronics Reliability* 13(4), 295-297.
- Brown, R.E., Burke, J.J., 2000. Managing the risk of performance based rates. *IEEE Transactions on Power Systems* 15(2), 893-898.
- Brown, R.E., Frimpong, G., Willis, H.L., 2004. Failure rate modeling using equipment inspection data. *Power Systems, IEEE Transactions on* 19(2), 782-787.
- Cohen, M.A., Agrawal, N., Agrawal, V., 2006. Winning in the aftermarket. *Harvard Business Review* 84(5), 129-138.
- Coit, D.W., 1997. System-reliability confidence-intervals for complex-systems with estimated component-reliability. *IEEE Transactions on Reliability* 46(4), 487-493.
- De Souza E Silva, E., Gail, H.R., 1986. Calculating cumulative operational time distributions of repairable computer systems. *IEEE Transactions on Computers* C-35(4), 322-332.
- Feizollahi, M.J., Modarres, M., 2012. The robust deviation redundancy allocation problem with interval component reliabilities. *IEEE Transactions on Reliability* 61(4), 957-965.
- Guajardo, J.A., Cohen, M.A., Kim, S.H., Netessine, S., 2012. Impact of performance-based contracting on product reliability: An empirical analysis. *Management Science* 58(5), 961-979.
- Van der Heijden, M., 1988. Interval uneffectiveness distribution for a k-out-of-n multistate reliability system with repair. *European Journal of Operational Research* 36(1), 66-77.
- Helander, A., Möller, K., 2007. System supplier's customer strategy. *Industrial Marketing Management* 36(6), 719-730.
- Huang, H.Z., Liu, Z.J., Murthy, D.N.P., 2007. Optimal reliability, warranty and price for new products. *IIE Transactions* 39(8), 819-827.
- Hussain, A.Z.M.O., Murthy, D.N.P., 1998. Warranty and redundancy design with uncertain quality. *IIE Transactions* 30(12), 1191-1199.
- Hussain, A.Z.M.O., Murthy, D.N.P., 2003. Warranty and optimal reliability improvement through product development. *IIE Transactions* 38(1113), 1211-1217.
- Hwang, C.L., 1975. Optimization of system reliability by the sequential unconstrained minimization technique. *Microelectronics Reliability* R24(2), 133-135.

- Jin, T.D., Coit, D.W., 2001. Variance of system reliability estimates with arbitrarily repeated components. *IEEE Transactions on Reliability* 50(4), 409-413.
- Jin, T.D., Tian, Y., 2012. Optimizing reliability and service parts logistics for a time-varying installed base. *European Journal of Operational Research* 218(1), 152-162.
- Jin, T.D., Tian, Y., Xie, M., 2015. A game-theoretical approach for optimizing maintenance, spares and service capacity in performance contracting. *International Journal of Production Economics* 161, 31-43.
- Jin, T.D., Wang, P., 2012. Planning performance based contracts considering reliability and uncertain system usage. *Journal of the Operational Research Society* 63(10), 1467-1478.
- Kim, S.H., Cohen, M.A., Netessine, S., 2007. Performance contracting in after-sales service supply chains. *Management Science* 53(12), 1843-1858.
- Kim, S.H., Cohen, M.A., Netessine, S., 2017. Reliability or inventory? an analysis of performance-based contracts for product support services, in: Ha, A., Tang, C. (Eds.), *Handbook of Information Exchange in Supply Chain Management*. Springer, pp. 65-88.
- Kuo, W., Wan, R., 2007. Recent advances in optimal reliability allocation. *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans* 37(2), 143-156.
- Li, D., Haimes, Y.Y., 1992. A decomposition method for optimization of large-system reliability. *IEEE Transactions on Reliability* 41(2), 183-188.
- Mettas, A., Kallenberg, R., 2000. Reliability allocation and optimization for complex systems, in: *Proceedings of the Annual Reliability and Maintainability Symposium*, Institute Electrical and Electronics Engineers, Los Angeles, USA.
- Minges, M.L., 1989. *Electronic Materials Handbook: Packaging, Volume I*. ASM International, Materials Park, USA.
- Monga, A., Zuo, M.J., 1998. Optimal system design considering maintenance and warranty. *Computers and Operations Research* 25(9), 691-705.
- Nakagawa, T., Goel, A.L., 1973. A note on availability for a finite interval. *IEEE Transactions on Reliability* 22(5), 271-272.
- Öner, K.B., Franssen, R., Kiesmüller, G., van Houtum, G.J., 2007. Life cycle costs measurement of complex systems manufactured by an engineer-to-order company, in: *The 17th International Conference on Flexible Automation and Intelligent Manufacturing*, Philadelphia, USA. pp. 1843-1858.

- Öner, K.B., Kiesmüller, G., van Houtum, G.J., 2010. Optimization of component reliability in the design phase of capital goods. *European Journal of Operational Research* 205(3), 615-624.
- Öner, K.B., Scheller-Wolf, A., van Houtum, G.J., 2013. Redundancy optimization for critical components in high-availability technical systems. *Operations Research* 61(1), 244-264.
- Park, M., Jung, K., Park, D.H., 2015. Optimal warranty policies considering repair service and replacement service under the manufacturer's perspective. *Annals of Operations Research*. To appear. doi:10.1007/s10479-014-1740-1.
- Ramirez-Marquez, J.E., Levitin, G., 2008. Algorithm for estimating reliability confidence bounds of multi-state systems. *Reliability Engineering and System Safety* 93(8), 1231-1243.
- Selçuk, B., Ağralı, S., 2013. Joint spare parts inventory and reliability decisions under a service constraint. *Journal of the Operational Research Society* 64(6), 446-458.
- Selviaridis, K., Wynstra, F., 2015. Performance-based contracting: a literature review and future research directions. *International Journal of Production Research* 53(12), 3505-3540.
- Takacs, L., 1957. On certain sojourn time problems in the theory of stochastic processes. *Acta Mathematica Hungarica* 8(1), 169-191.
- Tijms, H.C., 1994. *Stochastic models: an algorithmic approach*. Wiley, New York, USA.