Fleet readiness: stocking spare parts and high-tech assets

Abstract

We consider a maintenance shop that is responsible for the availability of a fleet of assets, e.g., trains. Unavailability of assets may be due to active maintenance time or unavailability of spare parts. Both spare assets and spare parts may be stocked in order to ensure a certain fleet readiness, which is the probability of having sufficient assets available for the primary process (e.g., running a train schedule) at any given moment. This is different from guaranteeing a certain average availability, as is typically done in the literature on spare parts inventories. We analyze the corresponding system, assuming continuous review and base stock control. We propose an algorithm, based on a marginal analysis approach, to solve the optimization problem of minimizing holding costs for spare assets and spare parts. Since the problem is not item separable, even marginal analysis is time consuming, but we show how to efficiently solve this. Using a numerical experiment, we show that our algorithm generally leads to a solution that is close to optimal, and that it is much faster than an existing algorithm for a closely related problem. We further show that the additional costs that are incurred when the problem of stocking spare assets and spare parts is not solved jointly, can be significant. A key managerial insight is that typically, the number of spare assets to be acquired is very close to a lower bound that is determined only by the active maintenance time on the assets. It is typically not cost effective to acquire more spare assets to cover spare parts unavailability.

Keywords: Maintenance · Inventory · Fleet sizing · Spare parts

1 Introduction

Many important services and operations depend on the availability of a sufficiently large fleet of assets. An airline, for example, depends on a fleet of aircraft to service all planned flights, while a railway company depends on a fleet of rolling stock to make the train schedule work. Other examples exist in the defense and maritime industries. In all such cases, the availability of assets (the fraction of time that they are available to operate) is not the most appropriate measure of fleet performance. For example, consider a train operator that requires 90 trains to run its schedule as planned, while it owns 100 trains that achieve a 90% availability. The train operator seems to do fine, but it could be that half of the time there are 95 trains available, while the other half of the time there are 85 trains available: during the latter half of the time, the train operator cannot run its schedule. Similarly, consider an air force that has brought a number of airplanes to a war zone. To fly a relatively safe mission, it requires 40 airplanes to be available. If the average availability of its 50 airplanes is 80%, the target seems to be achieved. However, it may be that most of the time about 35 airplanes are available, while sometimes

there are almost 50 available. Then very often, there are not enough airplanes available to fly a safe mission. In both examples, a more accurate measure of performance is the fraction of time that sufficient assets are available to fulfill the function of the fleet, i.e., the probability that sufficient assets are available at an arbitrary moment in time. We refer to this performance measure as fleet readiness.

When k assets are needed to fulfill the function of the fleet, then typically $N > k$ assets need to be acquired to achieve a high fleet readiness, since assets are subject to failures and need maintenance. The maintenance time of an asset consists of two main parts: the *active maintenance time* in which the actual maintenance operations occur (usually the replacement of line replaceable units) and the *maintenance delay time* which is the waiting time for maintenance resources to become available (some authors call this time to support). A major culprit for maintenance delay is a lack of spare parts needed for replacement.

High fleet readiness can be achieved by a combination of the following: (1) Buying assets in addition to what is necessary to run daily operations; (2) Reducing the maintenance delay time by stocking spare parts; (3) Reducing the required number of maintenance actions by increasing asset reliability; Or (4) improving the speed of maintenance/replacement operations. This paper focuses on the first two options as these amount to investment decisions of a logistical nature. The last two options can usually only be achieved by making asset engineering modifications that are specific to the technology of the asset.

Buying as many assets as a given budget allows is a popular method to increase fleet readiness but it is not always effective. The money needed to buy assets and spare parts usually comes from the same budget. In the last decades of the previous century, the Dutch defense engaged in what has come to be called 'carcass politics'[1]. Under carcass politics, the available budget to establish a fleet is spent as much as possible on buying complete assets, and the remainder is spent on spare parts. Spare parts become short in supply soon after this and as a result, technicians start using parts from complete assets leaving only a 'carcass' behind. This practice is often referred to as cannibalization. Clearly, this practice does not necessarily lead to high fleet readiness. There is a trade-off between investing in assets and spare parts to meet a certain fleet readiness and this paper explores this trade-off, using a model of a single stock point with continuous review, stochastic lead times, and base stock control.

The trade-off between investing in assets or spare parts to realize a certain fleet readiness objective is non-trivial. In general, this problem is non-convex and the analysis cannot be separated into an analysis per spare part type and asset: Its evaluation requires the convolution of backorder distributions per spare part type. This is in stark contrast with many spare part inventory problems in which the resulting optimization problems are convex and separable per item (see, e.g., Sherbrooke, 2004; Muckstadt, 2005; Basten and Van Houtum, 2014). Assets and spare parts achieve a certain fleet readiness jointly and so their analysis cannot be separated. In fact, we will show that their joint analysis is mathematically a generalization of multi-echelon inventory theory, even though we consider only a single stock point. Unfortunately,

[1]The Dutch word is 'rompenpolitiek', see, e.g., Tjepkema (2010)

this generalization is not susceptible to standard tools such as Clark and Scarf decomposition (Clark and Scarf, 1960) and METRIC type inventory models (Sherbrooke, 1968).

Our main contributions in this paper are the following: We consider the problem of deciding on asset investment and spare part investment jointly, whereas previous work considers them separately, see also Section 2, which we also often see in practice. However, both are sizeable investments that serve a common purpose: achieving high fleet readiness. Fleet readiness is usually not used as service measure in this setting because it is computationally demanding to evaluate and difficult to optimize. We show that evaluation requires performing a large number of convolutions, but we also show how the number of convolutions can be lowered exponentially. We next develop a greedy heuristic for this problem that is computationally efficient. In a numerical experiment, we compare our heuristic with enumeration on small instances and find that our heuristic finds the optimal solution of 51% of our test instances and has an average optimality gap on the other instances of 3.7%. Our algorithm is 50 times faster on medium size instances than an existing algorithm that was developed for a related problem. (The existing algorithm takes too much time to perform a comparison on large instances.) Our key managerial insight is that spare assets should be acquired to cover active maintenance time on the assets; it is typically not cost effective to acquire more spare assets to cover spare parts unavailability. In other words, the use of 'carcass politics' is not cost effective.

The remainder of this paper is organized as follows. We discuss related literature in Section 2 and position our work with respect to previous work. In Section 3, we explain the model and the optimization problem that we focus on. We analyze the model in Section 4; we show that the problem is not convex and we prove some other properties. We use those to construct an algorithm to solve the optimization problem in Section 5. In Section 6, we perform a numerical experiment, and we conclude in Section 7.

2 Related literature

Since our main contributions are the combination of the fleet sizing and spare part investment decisions subject to a service level constraint that is not often used, this literature review is structured as follows: We discuss the fleet readiness measure in Section 2.1, fleet sizing in Section 2.2 and spare parts optimization in Section 2.3. In Section 2.3, we specifically focus on a closely related paper by De Smidt-Destombes et al. (2011).

2.1 Fleet readiness

Fleet readiness as a performance measure is not as common as availability. Some authors however, already noted that in many instances readiness is a more appropriate performance measure. Safaei et al. (2011), for instance, consider a deterministic maintenance scheduling problem subject to a manpower constraint and a fleet readiness constraint. Jin and Wang (2012) use the fleet readiness measure in the context of performance based contracting. They approximate this measure by using the availability as the probability that a vehicle is available

at an arbitrary moment in time and then use the binomial distribution to compute the fleet readiness. This approximation is more tractable than actual fleet readiness but it assumes that the availability of different vehicles is uncorrelated at any particular time point. A similar approach has been followed by Costantino et al. (2013) in a multi-echelon, multi-indenture spare parts inventory setting. Some authors use the term fleet readiness as the average number of vehicles of a fleet that are available, e.g., Sherbrooke (1971) and Salman et al. (2007). That is, these authors consider the availability times the size of the fleet rather than the fleet readiness as we define it.

A closely related concept from the reliability engineering literature is the availability of a k-out-of-N system (e.g., De Smidt-Destombes et al., 2004). In this setting, a system consists of N components and only functions if at least k out of those N components are operational. The availability is then defined as the probability that at least k out of the N components are operational. In our setting, we would say that a fleet is ready if at least k out of N assets are operational, or alternatively, if not more than $N-k$ assets are unavailable. Thus these measures are equivalent.

2.2 Fleet sizing

Fleet sizing for vehicles has been studied in different settings. Hoff et al. (2010) and Pantuso et al. (2014) provide a review of these models in the general and maritime setting, respectively. Most of these models are deterministic and are concerned with calculating the minimum fleet size necessary to perform daily operations. Our model takes this minimum number of vehicles needed as an input and supports the investment decision in additional vehicles (or other assets) and spare parts to make sure that the fleet is operationally ready with a certain probability at any moment in time. Hoff et al. (2010) already note that dealing with uncertainty is an important aspect to incorporate when making the fleet sizing decision. Our work partially fills this gap by providing a model that deals with the uncertainty in the number of vehicles down for maintenance or lack of a spare part.

2.3 Spare parts optimization

The optimization of spare part inventory decisions has a long history that started with the work of Feeney and Sherbrooke (1966) and Sherbrooke (1968). This line of research has led to a large stream of literature that has been consolidated in the books by Sherbrooke (2004) and Muckstadt (2005) and the review papers by Guide Jr. and Srivastava (1997), Kennedy et al. (2002), and Basten and Van Houtum (2014). Here, we focus on the most closely related work, which is the paper of De Smidt-Destombes et al. (2011). In that paper, the authors consider a fleet that is taken on a mission with a package of spare parts. The objective is to minimize the investment in this spare parts package subject to a constraint on the probability that the fleet remains ready throughout the mission. We will show that that constraint is mathematically equivalent to the fleet readiness constraint as used in this paper. We extend the model in two ways: (1) We also consider the size of the fleet as a decision variable and (2) we account for

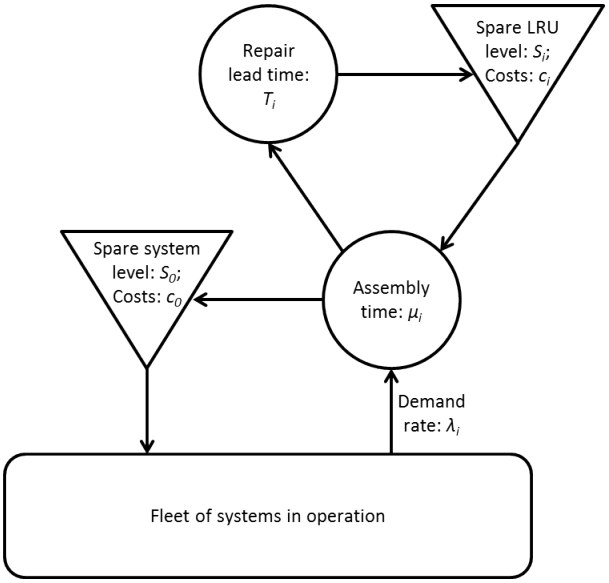

Figure 1: Modeled system

the fact that maintenance itself requires time and renders an asset unavailable. We show that, even for a fixed fleet size, optimizing the spare parts package is not a separable and convex problem. Despite this, De Smidt-Destombes et al. (2011) use a marginal analysis approach and we pursue a similar approach. As a new contribution, we benchmark this approach with respect to the optimal solution found by enumeration. We find that our algorithm yields high quality solutions. Furthermore, we provide results that make algorithms based on marginal analysis more tractable by giving easy to compute bounds so that gradients do not need to be computed for every direction of ascent. In addition, we provide an algorithm that computes the gradient in $O(\log n)$ time instead of $O(n)$ time, with n being the number of distinct spare part types.

3 Model description

The system that we analyze is shown in Figure 1. We consider a fleet of assets that are composed of line replaceable units (LRUs). We let I denote the set of LRUs and we reserve the index 0 for the assets; we denote the set of LRUs and assets by $I_0 = I \cup \{0\}$. Assets fail randomly due to a failure in exactly one LRU $i \in I$; such failures occur according to a Poisson process with intensity λ_i and the total intensity over all LRUs is denoted by $\lambda_0 = \sum_{i \in I} \lambda_i$. λ_i indicates the failure rate of LRU i over the complete fleet and it is constant over time, i.e., it is not influenced by the number of assets that is operational at any point in time. The assumption of a constant failure intensity is regularly made in the spare parts inventory literature, see, e.g., Basten and Van Houtum (2014, p.40). It is reasonable, since the required fleet readiness is typically high. Furthermore, when there is a shortage of operational assets, which would lead to a slightly lower failure intensity, then often some work load is routed to the operational assets so that their failure intensity increases.

An asset is repaired by replacement of the failed part by a functioning spare part. In the

remainder of this paper, if we refer to (spare) parts, components, or items of LRU type i, we say parts of LRU i. We assume that disassembly of the failed part takes negligible time (i.e., is instantaneous); assembly of the functioning spare part into the asset takes exactly μ_i time units for LRU $i \in I$ if a spare part is available immediately from stock (i.e., μ_i is deterministic). After being repaired, the asset is sent to the pool of stand-by assets. We also refer to this pool as the stock of spare assets.

The failed part of LRU $i \in I$ is sent to the repair shop; its repair lead time is generally distributed with mean T_i time units. Repair times of parts of the same LRU are independent and identically distributed (i.i.d.) and repair times of parts of different LRUs are independent of each other. In other words, we assume that the repair shop has an infinite number of servers, or that the repair shop is able to schedule repairs and hire capacity such that it can guarantee a certain average repair time (we have made an analogous assumption for the maintenance shop). After being repaired, a part is returned to stock. Repairs may be performed either at an internal repair shop, or they may be outsourced to an external repair shop. In fact, the model can also be used if parts are discarded and replaced by new parts. In that case, repair lead time should be read as supply lead time or order-and-ship time.

All stock points are controlled using a continuous review $(S_i - 1, S_i)$ base stock policy (i.e., one-for-one replenishment) with S_i being the base stock level for LRU $i \in I$ or the asset $(i = 0)$. (Notice that S_0 does not indicate the fleet size; if, say, 100 assets are required in the primary process of the user and the base stock level, S_0, is equal to 10, then the fleet size is 110.) Our assumptions mean that failed assets or parts are not batched, but immediately sent into maintenance or repair. If the setup costs for a repair are low, this is a reasonable assumption. As a result, this assumption is commonly made in the spare parts literature (see, e.g. Basten and Van Houtum, 2014, p.40); we come back to this assumption in Remark 3.1.

Under this policy, the dynamics of the system can be described as follows: Let $D_i(t', t)$ denote the demand for LRU i (or, equivalently, the number of failures in parts of LRU i) between times t' and t. Let $X_i(t)$ denote the number of parts of LRU $i \in I$ in repair, also called the pipeline of LRU i, at time t. Then, if the repair lead time T_i is deterministic, the pipeline at time t consists of all demands between times $t - T_i$ and t, i.e., $X_i(t) = D_i(t - T_i, t)$. Due to Palm's theorem (Palm, 1938), this equality still holds in distribution if the repair lead time is not deterministic. The number of backorders for LRU $i \in I$ is denoted by $B_i(t, S_i)$ and satisfies $B_i(t, S_i) = [X_i(t) - S_i]^+$, with $[x]^+ = \max\{0, x\}$. In other words, there are backorders if the number of parts in the pipeline is higher than the base stock level, with the number of backorders being equal to the difference. We denote by $Y_0(t)$ the number of assets in the maintenance shop that are actively being maintained at time t (i.e., the assets that are waiting for a spare part are not included). Since the assembly time μ_i is assumed to be deterministic, this is the summation over all LRUs of the demands for that LRU between times $t - \mu_i$ and t, i.e., $Y_0(t) = \sum_{i \in I} D_i(t - \mu_i, t)$. For notational convenience, we introduce $\mathbf{S}$ as the vector of all base stock levels S_i for $i \in I$. The pipeline $X_0(t, \mathbf{S})$ of assets in the maintenance shop at time t is equal to the number of assets that came in for maintenance after time $t - \mu_i$, plus the

number of assets that should have finished maintenance, but were delayed due to backordered spare parts:

$$X_0(t,\mathbf{S}) = Y_0(t) + \sum_{i\in I} B_i(t-\mu_i, S_i) = \sum_{i\in I} D_i(t-\mu_i, t) + \sum_{i\in I} \left[D_i(t-\mu_i-T_i, t-\mu_i) - S_i\right]^+, \quad (1)$$

while the number of assets short is denoted by $B_0(t,\mathbf{S_0}) = [X_0(t,\mathbf{S}) - S_0]^+$, with $\mathbf{S_0}$ being the vector of all base stock levels S_i for $i \in I_0$. (This can also be interpreted as the number of backordered assets.) The readiness, $R(\mathbf{S_0})$, is the probability of not being any assets short in steady state: $R(\mathbf{S_0}) = \lim_{t\to\infty} \mathbb{P}\{B_0(t,\mathbf{S_0}) = 0\}$. In other words, the readiness is the long-term probability that the number of assets in maintenance (i.e., in the pipeline) does not exceed the number of spare assets.

Remark 3.1. If the asset consists of one LRU only, our system simplifies to a two-echelon serial inventory system, the Clark-Scarf model (Clark and Scarf, 1960). Specifically, when $|I| = 1$, $Y_0(t)$ can be interpreted as the number of orders in transit from the upstream stock point to the downstream stock point at time t, while $B_1(t-\mu_1, S_1)$ represents the orders from the downstream stock point that are backordered at the upstream stock point at time $t-\mu$. By allowing $|I| > 1$, we are dealing with a generalization of a two-echelon serial inventory system under base stock control. Since base stock policies are optimal in the Clark-Scarf model and many of its generalizations, including some convergent systems, (see for an overview Van Houtum, 2006) our assumption of base stock control seems reasonable.

Remark 3.2. When $\mu_i = 0$ and $T_i = T$ for all $i \in I$, then $R(\mathbf{S_0})$ can also be interpreted as the probability that the fleet remains ready during a mission of length T when a spare parts package of size $\mathbf{S}$ is brought on the mission. (Note that $T_i = T$ for all $i \in I$ implies that spare parts cannot be repaired during the mission.) For a fixed asset base stock level S_0, this is the setting that De Smidt-Destombes et al. (2011) consider.

The costs of holding spare assets and LRUs are linear in their base stock level: c_i per unit for asset or LRU $i \in I_0$. Our goal is to find the base stock levels that minimize the total costs $C(\mathbf{S_0}) = \sum_{i\in I_0} c_i S_i$, such that the target readiness R^{obj} is achieved. Formally, our optimization problem, Problem (P), is:

$$(P) \qquad \min_{\mathbf{S_0}\in\mathbb{N}_0^{|I_0|}} \quad C(\mathbf{S_0})$$
$$\text{subject to} \quad R(\mathbf{S_0}) \geq R^{\text{obj}},$$

with $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ being the set of non-negative integers. We emphasize that Problem (P) is not separable per item because $R(\mathbf{S_0})$ cannot be written as a sum of terms that depend on one S_i only. Further note that the costs can also comprise initial investment costs instead of holding costs.

4 Analysis

In this section, we give results on the behavior of the fleet readiness as a function of the number of spare parts and spare assets. We use these results to explain, in Section 5, why we make certain choices in the algorithm that we use to solve Problem (P). Since we consider the system in steady state, we suppress the time parameter in the state variables from now on, and we show their distributions in Lemma 1. The results in this lemma are very similar to the derivation of results for the standard two-echelon model for spare parts (see, e.g., Graves, 1985; Basten and Van Houtum, 2013).

Lemma 1. *In steady state, the state variables are distributed as follows:*

(i) For $i \in I$, the pipeline, X_i, is Poisson distributed with mean $\lambda_i T_i$, i.e.:

$$\mathbb{P}\{X_i = x\} = \frac{(\lambda_i T_i)^x}{x!} e^{-\lambda_i T_i}, \quad x \in \mathbb{N}_0.$$

(ii) For $i \in I$, the distribution of the number of backorders, $B_i(S_i)$, is given by:

$$\mathbb{P}\{B_i(S_i) = b\} = \begin{cases} \mathbb{P}\{X_i \leq S_i\}, & \text{if } b = 0; \\ \mathbb{P}\{X_i = S_i + b\}, & \text{if } b \in \mathbb{N}. \end{cases}$$

(iii) The number of assets that are actively being maintained, Y_0, is Poisson distributed with mean $\sum_{i \in I} \lambda_i \mu_i$, i.e.:

$$\mathbb{P}\{Y_0 = y\} = \frac{\left(\sum_{i \in I} \lambda_i \mu_i\right)^y}{y!} e^{-\sum_{i \in I} \lambda_i \mu_i}, \quad y \in \mathbb{N}_0.$$

(iv) The distribution of the asset pipeline, $X_0(\mathbf{S})$, is given by:

$$\mathbb{P}\{X_0(\mathbf{S}) = x\} = \sum_{y=0}^{x} \mathbb{P}\{Y_0 = y\} \, \mathbb{P}\left\{\sum_{i \in I} B_i(S_i) = x - y\right\}, \quad x \in \mathbb{N}_0.$$

(v) The distribution of the number of assets short, $B_0(\mathbf{S_0})$, is given by:

$$\mathbb{P}\{B_0(\mathbf{S_0}) = b\} = \begin{cases} \mathbb{P}\{X_0(\mathbf{S}) \leq S_0\}, & \text{if } b = 0; \\ \mathbb{P}\{X_0(\mathbf{S}) = S_0 + b\}, & \text{if } b \in \mathbb{N}. \end{cases}$$

Proof. (i) Under a base stock policy, each demand for LRU $i \in I$ triggers an order, and each order is delivered after an i.i.d. amount of time with mean T_i. Therefore, the number of outstanding orders of LRU $i \in I$ behaves as the number of customers in an $M/G/\infty$-queue with arrival rate λ_i and processing time with mean T_i. By Palm's Theorem (Palm, 1938), the number of customers in such a queue in steady state is Poisson distributed with mean $\lambda_i T_i$.

(ii) The number of backorders at time t is defined as $B_i(t, S_i) = [X_i(t) - S_i]^+$. Therefore, we have $\mathbb{P}\{B_i(S_i) = 0\} = \mathbb{P}\{[X_i - S_i]^+ = 0\} = \mathbb{P}\{X_i \leq S_i\}$, and $\mathbb{P}\{B_i(S_i) = b\} = \mathbb{P}\{[X_i - S_i]^+ = b\} = \mathbb{P}\{X_i = S_i + b\}$ for $b \in \mathbb{N}$.

(iii) By definition, $Y_0(t) = \sum_{i \in I} D_i(t - \mu_i, t)$. Since μ_i is deterministic, $D_i(t - \mu_i, t)$ denotes the number of arrivals of a Poisson process with intensity λ_i in a time interval of length μ_i. Therefore $D_i(t - \mu_i, t)$ has a Poisson distribution with mean $\lambda_i \mu_i$. Since the Poisson distribution is closed under convolution, $Y_0(t)$ has a Poisson distribution with mean $\sum_{i \in I} \lambda_i \mu_i$ for each t and in particular as $t \to \infty$.

(iv) First observe that since μ_i is deterministic, $[t - \mu_i - T_i, t - \mu_i)$ and $[t - \mu_i, t)$ are disjoint intervals. By Equation (1) and the independent increments of the Poisson process, this implies that Y_0 and $\sum_{i \in I} B_i(S_i)$ are independent random variables. Now, using Equation (1) and letting $t \to \infty$, we have:

$$\begin{aligned}
\mathbb{P}\{X_0(\mathbf{S}) = x\} &= \mathbb{P}\left\{Y_0 + \textstyle\sum_{i \in I} B_i(S_i) = x\right\} \\
&= \sum_{y=0}^{x} \mathbb{P}\{Y_0 = y\}\, \mathbb{P}\left\{Y_0 + \textstyle\sum_{i \in I} B_i(S_i) = x \mid Y_0 = y\right\} \\
&= \sum_{y=0}^{x} \mathbb{P}\{Y_0 = y\}\, \mathbb{P}\left\{\textstyle\sum_{i \in I} B_i(S_i) = x - y\right\},
\end{aligned}$$

where the final equality follows from the independence of Y_0 and $\sum_{i \in I} B_i(S_i)$.

(v) This is analogous to part (ii).

□

We use additional notation in this section: Let $\mathbf{e}_i$ be a vector of length $|I_0|$ with all zeros, except at the location corresponding to the base stock level of spare assets ($i = 0$: S_0) or spare LRUs ($i \in I$: S_i). Furthermore, notice that concavity of $R(\mathbf{S_0})$ in S_i for $i \in I_0$ is equivalent to $R(\mathbf{S_0} + \mathbf{e}_i) - R(\mathbf{S_0}) \geq R(\mathbf{S_0} + 2\mathbf{e}_i) - R(\mathbf{S_0} + \mathbf{e}_i)$ for $\mathbf{S_0} \in \mathbb{N}_0^{|I_0|}$, while joint concavity in $i, j \in I_0$ is equivalent to $R(\mathbf{S_0} + \mathbf{e}_j) - R(\mathbf{S_0}) \geq R(\mathbf{S_0} + \mathbf{e}_i + \mathbf{e}_j) - R(\mathbf{S_0} + \mathbf{e}_i)$ for $\mathbf{S_0} \in \mathbb{N}_0^{|I_0|}$.

$R(\mathbf{S_0})$ is not in general jointly concave in S_0 and S_i with $i \in I$. As a counter example, consider an asset consisting of one LRU, indexed 1, with $\lambda_1 = 2$ and $\mu_1 = T_1 = 1$. Evaluating $R(\mathbf{S_0})$ gives the following results: $R(0,0) \approx 0.1353$, $R(1,0) \approx 0.4061$, $R(0,1) \approx 0.2707$, and $R(1,1) \approx 0.6090$. It is easily seen that if either S_0 or S_1 is increased, the readiness increases. However, $R(0,1) - R(0,0) \approx 0.1353 < R(1,1) - R(1,0) \approx 0.2030$, and $R(1,0) - R(0,0) \approx 0.2707 < R(1,1) - R(0,1) \approx 0.3383$. This means that $R(\mathbf{S_0})$ is not jointly concave. For larger values of S_0, the inequalities required for concavity do typically hold.

Because $R(\mathbf{S_0})$ is not jointly concave in general, our algorithm enumerates the number of spare assets (see Section 5); Proposition 1 gives bounds on its optimal value. The lower bound, S_0^{LB}, is the number of spare assets that are required to achieve the target readiness if there are unlimited spare parts available. In Section 6, we show that this lower bound is usually tight,

implying that spare assets are mainly stocked to cover the active maintenance time, while they are not stocked to cover spare parts unavailability.

Proposition 1. *The optimal number of spare assets for Problem (P), denoted as S_0^*, is bounded as follows:*

(i) $S_0^* \geq S_0^{LB}$, with S_0^{LB} being the smallest integer S that satisfies $\mathbb{P}\{Y_0 \leq S\} \geq R^{obj}$.

(ii) $S_0^* \leq S_0^{UB}$, with S_0^{UB} being the smallest integer S for which there exists $\mathbf{S_0}' = \left(S_0', S_1', \ldots, S_{|I|}'\right)$ with $S_0^{LB} \leq S_0' \leq S$, $C(\mathbf{S_0}') < c_0(S+1)$ and $R(\mathbf{S_0}') \geq R^{obj}$.

Proof. For part (i): Since $X_0(\mathbf{S}) \stackrel{d}{=} Y_0 + \sum_{i \in I} B_i(S_i)$ by definition ($\stackrel{d}{=}$ denotes equality in distribution), we have that $\mathbb{P}\{X_0(\mathbf{S}) \geq x\} \geq \mathbb{P}\{Y_0 \geq x\}$ for all $x \in \mathbb{N}_0$ because $B_i(S_i)$ are non-negative random variables. The readiness constraint in Problem (P) requires $\mathbb{P}\{X_0(\mathbf{S}) \leq S_0\} \geq R^{\text{obj}}$, so a feasible S_0 must satisfy $\mathbb{P}\{Y_0 \leq S_0\} \geq R^{\text{obj}}$.

For part (ii): $\mathbf{S_0}'$ represents a feasible solution, since $R(\mathbf{S_0}') \geq R^{\text{obj}}$, and the associated cost of this solution is $C(\mathbf{S_0}')$. Stocking $S+1$ spare assets costs $c_0(S+1)$. If $c_0(S+1) > C(\mathbf{S_0}')$, then the optimal solution cannot contain $S+1$ spare assets (or more), because there is a feasible solution with cost lower than $c_0(S+1)$. This means that S is an upper bound on the number of spare assets in the optimal solution, S_0^*. □

Problem (P) is also not in general jointly concave in S_i and S_j with $i, j \in I$ and $i \neq j$. As a counter example, consider an asset consisting of two LRUs, indexed 1 and 2, with $S_0 = \mu_1 = \mu_2 = 0$. Joint concavity is equivalent to $R(0, \mathbf{S}+\mathbf{e}_1) - R(0, \mathbf{S}) \geq R(0, \mathbf{S}+\mathbf{e}_1+\mathbf{e}_2) - R(0, \mathbf{S}+\mathbf{e}_2)$ for $\mathbf{S_0} \in \mathbb{N}_0^{|I_0|}$, which would mean that

$$
\begin{aligned}
&(1 - \mathbb{P}\{X_1 \leq S_1 + 1\}\,\mathbb{P}\{X_2 \leq S_2\}) - (1 - \mathbb{P}\{X_1 \leq S_1\}\,\mathbb{P}\{X_2 \leq S_2\}) \\
&\qquad \geq (1 - \mathbb{P}\{X_1 \leq S_1 + 1\}\,\mathbb{P}\{X_2 \leq S_2 + 1\}) - (1 - \mathbb{P}\{X_1 \leq S_1\}\,\mathbb{P}\{X_2 \leq S_2 + 1\}), \\
&\mathbb{P}\{X_1 \leq S_1\}\,\mathbb{P}\{X_2 \leq S_2\} - \mathbb{P}\{X_1 \leq S_1 + 1\}\,\mathbb{P}\{X_2 \leq S_2\} \\
&\qquad \geq \mathbb{P}\{X_1 \leq S_1\}\,\mathbb{P}\{X_2 \leq S_2 + 1\} - \mathbb{P}\{X_1 \leq S_1 + 1\}\,\mathbb{P}\{X_2 \leq S_2 + 1\}, \\
&(\mathbb{P}\{X_1 \leq S_1\} - \mathbb{P}\{X_1 \leq S_1 + 1\})\mathbb{P}\{X_2 \leq S_2\} \\
&\qquad \geq (\mathbb{P}\{X_1 \leq S_1\} - \mathbb{P}\{X_1 \leq S_1 + 1\})\mathbb{P}\{X_2 \leq S_2 + 1\}, \text{ and} \\
&\mathbb{P}\{X_2 \leq S_2\} \geq \mathbb{P}\{X_2 \leq S_2 + 1\}.
\end{aligned}
$$

However, $\mathbb{P}\{X_2 = S_2 + 1\} > 0$ for all $S_2 \geq 0$, so that $\mathbb{P}\{X_2 \leq S_2 + 1\} > \mathbb{P}\{X_2 \leq S_2\}$, showing that this problem is not jointly concave.

We show other convexity results in Proposition 2, which our algorithm uses to determine lower bounds. The proof can be found in Appendix A; it uses properties of the shape of the Poisson distribution.

Proposition 2. *The second order difference function for the number of spare LRUs $i \in I$, $\Delta_i^2 R(\mathbf{S_0}) = \Delta_i R(\mathbf{S_0} + \mathbf{e}_i) - \Delta_i R(\mathbf{S_0})$, behaves as follows:*

(i) If $S_0 + S_i < \lceil \lambda_i T_i \rceil - 2$, then $\Delta_i^2 R(\mathbf{S_0}) > 0$, that is, $R(\mathbf{S_0})$ is strictly convex in S_i.

(ii) If $S_i \geq \lceil \lambda_i T_i \rceil - 2$*, then* $\Delta_i^2 R(\mathbf{S_0}) \leq 0$*, that is,* $R(\mathbf{S_0})$ *is concave in* S_i*.*

Notice that:

- A similar result as part (ii) has been shown by Rustenburg (2000, p.41).
- The behavior of $\Delta_i^2 R(\mathbf{S_0})$ is not clear beforehand in all cases that are not covered by Proposition 2 (i.e., if both $S_i < \lceil \lambda_i T_i \rceil - 2$ and $S_0 + S_i \geq \lceil \lambda_i T_i \rceil - 2$).

Proposition 3 gives a result that allows our algorithm to avoid performing unnecessary calculations: Our algorithm uses a marginal analysis approach. In each iteration, an additional spare part is stocked of the LRU that gives the biggest 'bang for the buck'. Proposition 3 gives an upper bound on how much this 'bang for the buck' may have changed for a certain LRU from one iteration to the next. Our algorithm can thus quickly check if the 'bang for the buck' of a certain LRU may be sufficiently high to perform time consuming exact calculations. Since the proof is long and does not give insight into the problem, it is deferred to Appendix A.

Proposition 3. *If* $S_i \geq \lceil \lambda_i T_i \rceil - 2$ *and* $S_j \geq \lceil \lambda_j T_j \rceil - 2$*, with* $i, j \in I$*, then:*

$$\Delta_i R(\mathbf{S_0} + \mathbf{e}_j) - \Delta_i R(\mathbf{S_0}) < \mathbb{P}\{X_j = S_j + 1\}\, \mathbb{P}\{X_i = S_i + 1\}.$$

5 Algorithm

We give the pseudo code of our algorithm in Figure 2 and we explain the complete algorithm in Section 5.1. Next, we focus on how to compute the convolutions in Line 11 of our algorithm in Section 5.2. This is a very time consuming step in the algorithm and we propose a novel way to do this efficiently.

5.1 Overview

The algorithm functions as follows. It enumerates the asset base stock level between a lower bound (Line 1) and an upper bound (Line 3), based on parts (i) and (ii) of Proposition 1, respectively. Although those bounds are simple, we still find in our numerical experiment (Table 7) that typically, the number of enumerated spare asset levels is small, i.e., the difference between the lower and upper bounds is small.

For each asset base stock level, each LRU base stock level is initialized at a lower bound based on part (ii) of Proposition 2 (Line 4). Notice that this lower bound guarantees that the readiness is concave in each LRU base stock level. However, notice further that it is not guaranteed that the optimal LRU base stock level is above this lower bound: Consider a system in which all LRUs are inexpensive, but there is one very expensive LRU i with a failure rate λ_i and repair time T_i such that $\lambda_i T_i = 2 + \epsilon$, with ϵ being a very small number. The lower bound then ensures that $S_i = 1$, while it may be optimal to have $S_i = 0$. In practice, however, such examples are seldom encountered and each optimal LRU base stock level will typically be above the lower bound.

1: $S_0 \leftarrow \min\{S \in \mathbb{N}_0 \mid \mathbb{P}\{Y_0 \leq S\} \geq R^{\text{obj}}\}$
2: Calculate the probability mass function of Y_0
3: **while** $c_0 S_0 \leq C^{\text{best}}$ **do**
4: $S_i \leftarrow \max\{0, \lceil \lambda_i T_i \rceil - 2\}$ for all $i \in I$
5: Calculate the probability mass functions of B_i for all $i \in I$, and of $Y_0 + \sum_{i \in I} B_i$
6: $R^{\text{cur}} \leftarrow R(\mathbf{S_0})$; $\Gamma^{\text{best}} \leftarrow 0$
7: $i^{\text{best}} \leftarrow -1$; $\mathbb{P}\{X_{-1} = S_{-1}\} \leftarrow 1$; $\Gamma_i \leftarrow 1/c_i$ for all $i \in I$
8: **while** $R^{\text{cur}} < R^{\text{obj}}$ **do**
9: **for** $i \in I$ **do**
10: $\Gamma_i \leftarrow \Gamma_i + \frac{\mathbb{P}\{X_{i^{\text{best}}} = S_{i^{\text{best}}}\}\mathbb{P}\{X_i = S_i + 1\}}{c_i}$
11: **if** $\Gamma_i \geq \Gamma^{\text{best}}$ or $i = i^{\text{best}}$ **then**
12: $\Gamma_i \leftarrow \frac{R(\mathbf{S_0} + \mathbf{e}_i) - R^{\text{cur}}}{c_i}$
13: $\Gamma^{\text{best}} \leftarrow \max\{\Gamma_i, \Gamma^{\text{best}}\}$
14: **end if**
15: **end for**
16: $i^{\text{best}} \leftarrow \arg\max_{i \in I} \Gamma_i$; $S_{i^{\text{best}}} \leftarrow S_{i^{\text{best}}} + 1$; $R^{\text{cur}} \leftarrow R(\mathbf{S_0})$
17: **end while**
18: **if** $C(\mathbf{S_0}) < C^{\text{best}}$ **then**
19: $C^{\text{best}} \leftarrow C(\mathbf{S_0})$
20: **end if**
21: $S_0 \leftarrow S_0 + 1$
22: **end while**

Figure 2: Greedy algorithm for Problem (P)

Then, the probability mass functions are computed of B_i for $i \in I$, and of $Y_0 + \sum_{i \in I} B_i$ (Line 5). Using a smart way of ordering and storing these computations results in a reduction of the computations that are performed per iteration of the marginal analysis approach that is used to stock additional spare parts (Line 12). We explain this in detail in Section 5.2. Further variables are initialized in Lines 6 and 7.

As long as the target readiness has not been reached (Line 8), for each LRU (Line 9) an upper bound on the increase in readiness per additionally invested dollar is calculated (Line 10), using the result in Proposition 3. If the upper bound is such that the current LRU would be the best option encountered thus far (Line 11), exact calculations are performed (Line 12) and it is checked if it is really the best LRU encountered thus far (Line 13). In our numerical experiment (Table 6), we find that using the upper bound for a first check, saves over 50% of computation time for problem instances with 256 LRUs and that the relative savings increase with an increasing problem size.

Notice that Line 7 ensures that in the first iteration of the while loop (Lines 8 to 17), the first condition of the if-clause on Line 11 is always true. This means that the first LRU that is checked, is by definition the best encountered thus far. The second condition of that if-clause is required because Proposition 3, which is used in Line 10, applies only for $i \neq j$.

In Line 16, the base stock level of the right LRU is increased and the readiness is adapted. As soon as the target readiness is reached for the current asset base stock level (Line 8), the

marginal analysis approach is stopped and it is checked if a new best solution is found (Line 18). Next, the asset base stock level is increased by one (Line 21) so that a new iteration can start if the upper bound has not been reached yet (Line 3).

5.2 Convolutions

The computationally most demanding step in Algorithm 2 is the computation of $\Gamma_i = (R(\mathbf{S}_0 + \mathbf{e}_i) - R^{\text{cur}})/c_i$ (Line 12), specifically the evaluation of $R(\mathbf{S}_0 + \mathbf{e}_i) = \mathbb{P}\{B_0(\mathbf{S_0} + \mathbf{e}_i) = 0\} = \mathbb{P}\{U(\mathbf{S} + \mathbf{e}_i) \leq S_0\}$, with $U(\mathbf{S}) = Y_0 + \sum_{i \in I} B_i(S_i)$. This requires computing the probability mass function of $U(\mathbf{S} + \mathbf{e}_i)$ by convolution; we provide an algorithm to compute that using results that have already been computed for $U(\mathbf{S})$.

We require some additional notation. Let $\mathbf{B}_i(S_i)$ be a vector containing the probability mass function of $B_i(S_i) = (X_i - S_i)^+$ up to S_0, i.e., $\mathbf{B}_i(S_i) = (\mathbb{P}\{B_i(S_i) = 0\}, \ldots, \mathbb{P}\{B_i(S_i) = S_0\})$. Similarly, let $\mathbf{Y}_0 = (\mathbb{P}\{Y_0 = 0\}, \ldots, \mathbb{P}\{Y_0 = S_0\})$ and $\mathbf{U}(\mathbf{S}) = (\mathbb{P}\{U(\mathbf{S}) = 0\}, \ldots, \mathbb{P}\{U(\mathbf{S}) = S_0\})$. Furthermore, let $\mathbf{a} * \mathbf{b}$ denote the convolution of the vectors $\mathbf{a}$ and $\mathbf{b}$ of equal length: If $\mathbf{c} = \mathbf{a} * \mathbf{b}$, then $\mathbf{c}$ has the same length as both $\mathbf{a}$ and $\mathbf{b}$ and the i-th element of $\mathbf{c}$ is given by $c_i = \sum_{j=0}^{i} a_{i-j} b_j$. (Note that we number elements in a vector starting from 0.) The convolution operator satisfies commutativity ($\mathbf{a} * \mathbf{b} = \mathbf{b} * \mathbf{a}$) and associativity ($(\mathbf{a} * \mathbf{b}) * \mathbf{c} = \mathbf{a} * (\mathbf{b} * \mathbf{c})$). Finally, we let $\mathbf{B}_{a,b}(\mathbf{S}) = \mathbf{B}_a(S_a) * \cdots * \mathbf{B}_b(S_b)$ for $a \leq b$.

The complexity in computing $\mathbf{U}(\mathbf{S})$ lies in the computation of $\mathbf{B}_{1,|I|}(\mathbf{S})$ and its complexity increases with $|I|$. After computing $\mathbf{B}_i(S_i)$ for all $i \in I$, the straightforward way to compute $\mathbf{B}_{1,|I|}(\mathbf{S})$ is to successively compute as follows: $\mathbf{B}_{1,2}(\mathbf{S}) = \mathbf{B}_1(S_1) * \mathbf{B}_2(S_2)$, $\mathbf{B}_{1,3}(\mathbf{S}) = \mathbf{B}_{1,2}(\mathbf{S}) * \mathbf{B}_3(S_3)$, …, $\mathbf{B}_{1,|I|}(\mathbf{S}) = \mathbf{B}_{1,|I|-1}(\mathbf{S}) * \mathbf{B}_{|I|}(S_{|I|})$. This requires performing $|I| - 1$ convolutions and this is what the algorithm of De Smidt-Destombes et al. (2011) does.

Instead, our algorithm builds up a tree starting from its leaves. An example for $|I| = 8$ is shown in Figure 3. Formally the procedure works as follows: Compute $\mathbf{B}_{1,2}(\mathbf{S}) = \mathbf{B}_1(S_1) * \mathbf{B}_2(S_2)$, …, $\mathbf{B}_{|I|-1,|I|}(\mathbf{S}) = \mathbf{B}_{|I|-1}(S_{|I|-1}) * \mathbf{B}_{|I|}(S_{|I|})$. Then, compute $\mathbf{B}_{1,4}(\mathbf{S}) = \mathbf{B}_{1,2}(\mathbf{S}) * \mathbf{B}_{3,4}(\mathbf{S})$, …, $\mathbf{B}_{|I-3|,|I|}(\mathbf{S}) = \mathbf{B}_{|I|-3,|I|-2}(\mathbf{S}) * \mathbf{B}_{|I|-1,|I|}(\mathbf{S})$. Continue in this manner until arriving at the root node of the tree: $\mathbf{B}_{1,|I|}(\mathbf{S})$. This procedure also requires $|I| - 1$ convolutions. However, computing $\mathbf{B}_{1,|I|}(\mathbf{S} + \mathbf{e}_i)$, for some $i \in I$, can now be done efficiently by reusing most results in the tree: $\mathbf{B}_{a,b}(\mathbf{S} + \mathbf{e}_i) = \mathbf{B}_{a,b}(\mathbf{S})$ whenever $i < a$ or $b < i$. This is easily seen when we reconsider the example where $|I| = 8$: Figure 4 shows the tree that results when computing $\mathbf{B}_{1,8}(\mathbf{S} + \mathbf{e}_3)$; if $\mathbf{B}_{1,8}(\mathbf{S})$ is already evaluated, this only requires the evaluation of 4 nodes. Of those 4 nodes, one concerns the determination of $\mathbf{B}_3(S_3 + 1)$ and $3 = \log_2(8)$ require taking a convolution. The same reasoning can be applied for general $|I|$ and yields the following result.

Proposition 4. *After an initial evaluation of* $\mathbf{B}_{1,|I|}(\mathbf{S})$ *which requires* $O(|I|)$ *convolutions, all subsequent evaluations of* $\mathbf{B}_{1,|I|}(\mathbf{S}+\mathbf{e}_i)$ *with* $i \in I$ *require performing only* $O(\log |I|)$ *convolutions.*

The only thing that we have not explained yet is when to perform the convolution with Y_0. It would be straightforward to do this at the end (i.e., at the root of the tree), but that would require an additional convolution each time that an LRU stock level is increased. Therefore,

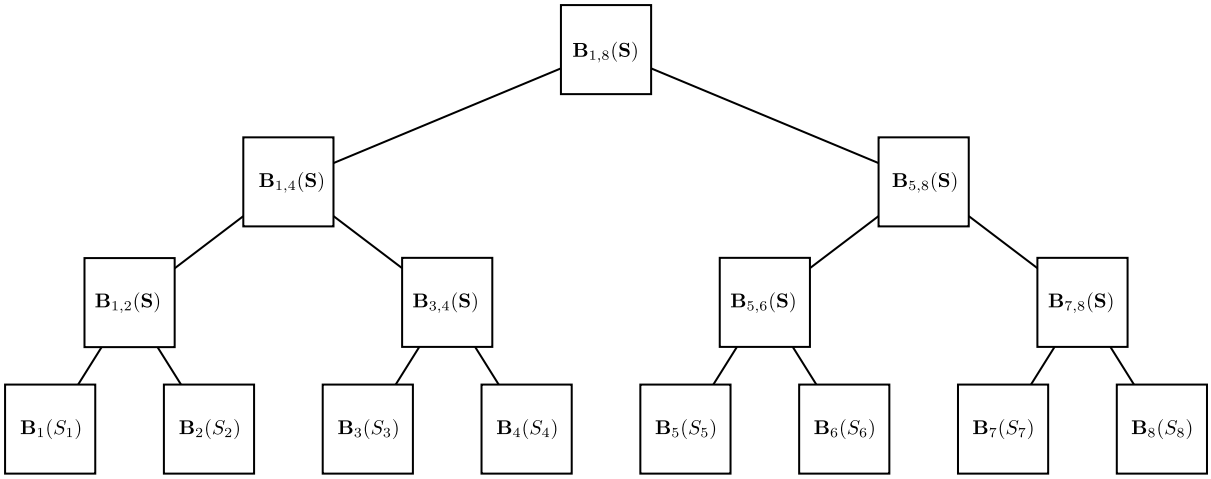

Figure 3: Computation of $\mathbf{B}_{1,|I|}(\mathbf{S})$ for $|I| = 8$ via a tree structure. Each non-leaf node in this tree is obtained by convolution of its two children nodes.

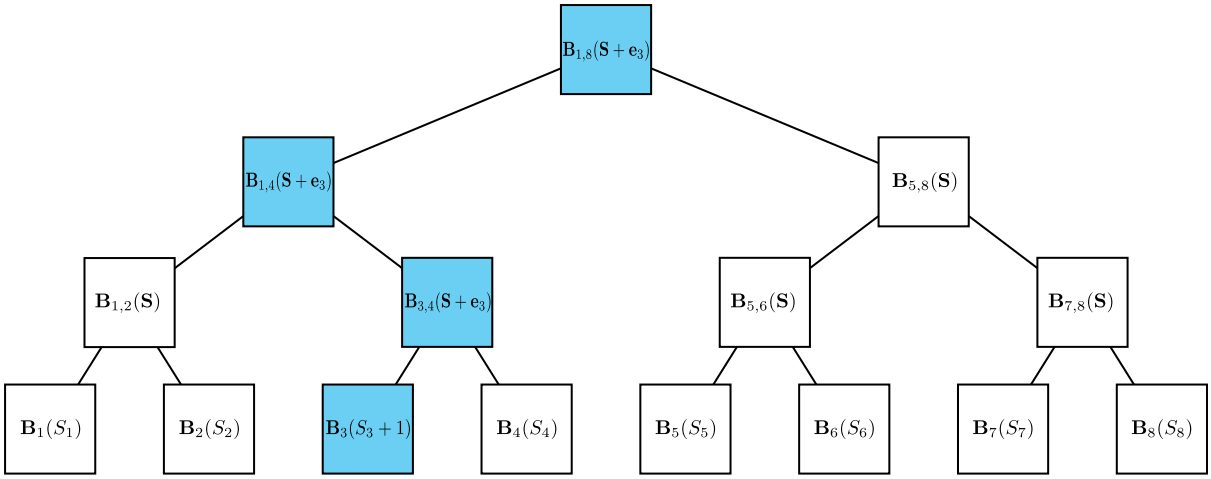

Figure 4: Computation of $\mathbf{B}_{1,|I|}(\mathbf{S}+\mathbf{e}_3)$ for $|I| = 8$ via a tree structure. This tree is identical to the tree for the computation of $\mathbf{B}_{1,|I|}(\mathbf{S})$ except in the shaded nodes.

this convolution is performed in the beginning: First, $\mathbf{B}_{1,1}(\mathbf{S}) = \mathbf{Y}_0 * \mathbf{B}_1(S_1)$ is calculated, which is used to calculate $\mathbf{B}_{1,2}(\mathbf{S}) = \mathbf{B}_{1,1}(\mathbf{S}) * \mathbf{B}_2(S_2)$.

6 Numerical experiment

We use the numerical experiment to answer four questions:

1. What is the quality of the solutions of our algorithm?
2. How should investments be divided between assets and spare parts?
3. What is the value of jointly optimizing spare assets and spare parts?
4. What is the computational effort required to run our algorithm?

		Set 1	Set 2
# LRUs	$\|I\|$	2; 4; 8	16; 64; 256; 1,024

		Set 1 & Set 2
Maximum assembly time	$\mu^{\max}$	0.001; 0.01
Maximum resupply lead time	$T^{\max}$	0.01; 0.1
Average costs of an LRU	c^{ave}	100; 1,000
Relative costs of an asset	c^{rel}	0.5; 1; 2
Target readiness	R^{obj}	0.9; 0.95; 0.975

Table 1: Settings of the parameters that are varied in the numerical experiment

To this end, we generate two sets of problem instances. Set 1 consists of smaller problem instances and is mainly used to answer Question 1, by comparing the solution of our algorithm with the optimal solution, found by enumeration. Set 2 consists of larger problem instances and is used for two reasons. First, it is used to get managerial insights, i.e., to answer Questions 2 and 3. Second, it is used to compare the computation times of our algorithm with that of De Smidt-Destombes et al. (2011), thus answering Question 4. We explain how we generate both sets of problem instances in Section 6.1. We answer Question 1 in Section 6.2, Question 2 and 3 in Section 6.3, and Question 4 in Section 6.4.[2]

6.1 Setup

Table 1 shows the settings for the parameters that are varied in our numerical experiment for the two sets of problem instances. (When the number of LRUs is varied, also λ_i is varied; we explain this below.) We use a full factorial design per set and we generate ten problem instances per combination of parameters. As a result, Set 1 and Set 2 consist of 2,160 and 2,880 problem instances, respectively. The way in which we generate problem instances leads to instances that are realistic in practice, and to a wide range of combinations of parameter values in the sets.

There are $|I|$ LRUs and a value μ is drawn from a uniform distribution on the range $[0, \mu^{\max}]$. Then, for each LRU $i \in I$ (see the explanation below):

- $\mu_i \leftarrow \mu$,
- T_i is drawn from a uniform distribution on the range $[0, T^{\max}]$,
- $\lambda_i \leftarrow \frac{128}{|I|}$ in Set 1 and $\lambda_i \leftarrow \frac{1{,}024}{|I|}$ in Set 2, and
- c_i is drawn from an exponential distribution with mean $\frac{1}{c^{\text{ave}}}$, and 10 is added. This effectively means that there are no LRUs with costs of less than 10, and the mean costs are $c^{\text{ave}} + 10$.

[2]The experiment is implemented in Python 3.4 and performed on an Intel Xeon E5530 @ 2.4 GHz with 8 GB RAM, running Windows Server 2008 R2 Enterprise Service Pack 1.

# LRUs: $\|I\|$	2	4	8
% Problem instances with optimal solution	73%	55%	26%
Average additional costs in remaining instances	2.8%	3.8%	4.0%
Maximum additional costs in remaining instances	63%	40%	93%

Table 2: Set 1: Quality of solutions found by our algorithm

The same value μ can be used for all $i \in I$, since it influences only the number of assets in active maintenance, Y_0.

λ_i is relevant only for calculating the distribution of the number of parts in resupply, X_i for $i \in I$, and the distribution of the number of assets in active maintenance, Y_0. Since the average number of parts in resupply is varied by varying T_i and since the average number of assets in the maintenance shop is varied by varying μ, λ_i can be kept constant in each problem instance. However, when the number of LRUs is varied, λ_i is also varied. Our aim is to get solutions in which the optimal number of spare assets and spare parts is realistic and higher than zero. We therefore show the average number of spare assets and spare parts in the solutions in the next section. The largest problem instances of Set 2 are the most realistic ones, with 1,024 LRUs and a demand rate per LRU of 1.

The costs of holding a spare asset are equal to c^{rel} times the summation of the costs of holding one of each of the spare parts, i.e., $c_0 = c^{\text{rel}} \sum_{i \in I} c_i$. Finally, the target readiness, R^{obj}, is varied.

6.2 Quality of solutions

Whenever we give results for multiple parameter settings in any of the tables in this or the next sections, then the numbers are the averages over all problem instances with one of those settings. We sometimes additionally give a line with a maximum, which is then the maximum over all problem instances with one of those settings. For example, in Table 3, the first value on the top line gives the average number of spare assets in the solutions for the problem instances with 16 LRUs, while the second value on that line gives the maximum number of spare assets in any of the solutions to the problem instances with 16 LRUs.

We use Table 2 to answer Question 1: It shows for the problem instances in Set 1 how our algorithm performs compared with the optimal solution. Many problem instances, 51%, are solved to optimality by our marginal analysis approach, and the average difference with the optimal solution on the other instances is small: 3.7% on average. (The values vary depending on the problem size.) The maximum difference is large, 93%, but large differences for these small instances can be caused by small mistakes. For example, for each of the three problem instances on which our algorithm incurs more than 50% additional costs, the number of stocked spare parts is correct, but one spare part is of the wrong LRU type. Furthermore, for all but one of the fourteen problem instances on which our algorithm incurs more than 25% additional costs, the achieved readiness is at least two percentage points higher, i.e., there is a large overshoot (remember that the readiness target is at least 90% in our problem instances). The

		# Spare assets in solution	Maximum	# Spare assets above the LB	Maximum	# Spare parts in solution	Divided by # LRUs	Additional costs of LB algorithm (%)	Maximum (%)
# LRUs: $\|I\|$	16	5.6	18	0.59	4	67	4.2	3.9	85
	64	5.2	17	0.20	1	151	2.4	2.1	76
	256	5.1	17	0.09	1	396	1.5	1.0	55
	1,024	5.0	17	0.04	1	1,251	1.2	0.5	43
Relative costs of an asset: c^{rel}	0.5	5.6	18	0.55	4	445	2.1	4.4	85
	1	5.1	17	0.12	1	470	2.4	1.1	46
	2	5.0	17	0.02	1	483	2.5	0.2	24
Maximum assembly time: $\mu^{\max}$	0.001	1.9	6	0.27	3	487	2.4	3.0	85
	0.01	8.5	18	0.19	4	445	2.2	0.8	34
Maximum resupply lead time: $T^{\max}$	0.01	5.1	18	0.11	2	300	1.1	1.3	85
	0.1	5.3	18	0.35	4	632	3.5	2.4	62
Average costs of an LRU: c^{ave}	100	5.2	18	0.23	3	461	2.3	1.9	85
	1000	5.2	18	0.23	4	471	2.4	1.9	84
Target readiness: R^{obj}	0.9	4.6	16	0.20	4	449	2.2	1.1	34
	0.95	5.2	17	0.22	3	467	2.3	1.6	51
	0.975	5.8	18	0.27	3	482	2.4	2.9	85

Table 3: Set 2: Key results for all parameters

only exception is one problem instance with four LRUs in which our algorithm stocks one asset too many. If we look at the problem instances for which our algorithm achieves a lower readiness than in the optimal solution, we see that the highest additional costs are 13% (the difference in readiness is 0.2 percentage point). All in all, we believe that our algorithm typically finds good solutions, especially considering that the overshoot typically decreases if the problem size increases (see, e.g. Basten and Van Houtum, 2014, p.42).

6.3 Managerial insights

To answer Question 2, Table 3 gives insight into the solutions that our algorithm finds (we explain the lower bound algorithm below). Recall that the lower bound (LB) in part (i) of Proposition 1, the lower bound that our algorithm uses, is such that the spare assets cover

# Spare assets above the lower bound	0	1	2	3	4	> 4
# Problem instances	2,349	424	83	23	1	0

Table 4: Set 2: Analysis of spare asset levels

# LRUs: $\|I\|$	2	4	8
# Spare assets in optimal solution	2.8	1.8	1.6
– maximum	12	6	5
# Spare assets above the lower bound	1.5	0.6	0.4
– maximum	11	5	3
# Spare parts in optimal solution	5.2	7.6	12.3
– divided by # LRUs	2.6	1.9	1.5

Table 5: Set 1: Analysis of optimal solutions

the active maintenance time. That means that if our algorithm stocks more spare assets, then apparently, spare assets are also used to cover spare parts unavailability. However, this is not the case in our numerical experiment: The number of spare assets in the solution is close to the lower bound; it is the same for 82% of the problem instances, see Table 4. Furthermore, the gap becomes smaller when the problem size increases, to 0.04 on average for problem instances with 1,024 LRUs. In fact, for more than 16 LRUs, the gap is never more than 1. We further see that a gap of more than 1 only occurs if assets are relatively inexpensive, i.e., $c^{\text{rel}} = 0.5$. However, this value of c^{rel} will not be realistic in most practical settings. The other parameters have a very limited influence on the number of additional spare assets to stock. Table 5 shows that also in the optimal solution, on Set 1, the number of spare assets is typically close to the lower bound.

This suggests that the lower bound is useful in practice to get an idea of the fleet size to acquire, while it is easy to calculate. For that reason, we have also implemented what we call the lower bound algorithm. Our algorithm, as defined in Section 5 and Figure 2, stocks spare parts for a number of spare asset levels, starting at the lower bound in part (i) of Proposition 1. To show the value of jointly optimizing the spare asset and spare parts stock levels, we need to compare the solutions of our algorithm with those of an algorithm in which first the spare asset level is determined and then the spare parts stock levels. For a fair comparison, the spare asset level should be determined in a meaningful way. Since we have seen above that our algorithm often finds a solution in which the lower bound on the asset stock level is used, this is the asset stock level that the lower bound algorithm uses. So, the lower bound algorithm first sets the number of spare assets to stock equal to the lower bound from part (i) of Proposition 1, and then stocks spare parts in the same way as our algorithm does. By comparing solutions of our algorithm with those of the lower bound algorithm, we answer Question 3.

Table 3 shows the additional costs that the lower bound algorithm incurs for the different parameter settings. We see that the additional costs are sometimes large. Although the average

# LRUs: $\|I\|$	16	64	256
(1) Computation time (seconds): using the bound	0.4	3.6	22.6
(2) Computation time (seconds): not using the bound	0.5	5.5	48.9
(3) Computation time (seconds): De Smidt-Destombes et al.	1.4	41.8	1,146.8
Relative computation time (2)/(1)	1.2	1.5	2.2
Relative computation time (3)/(1)	3.3	11.6	50.7
Relative computation time (3)/(2)	2.7	7.6	23.5

Table 6: Set 2: Computation times of our algorithms and that of De Smidt-Destombes et al. (2011), with 'bound' referring to the use (or not) of the results in Proposition 3

additional costs decreases when the number of LRUs increases, there is still an instance with 1,024 LRUs where using the lower bound algorithm leads to 43% additional costs. The lowest average additional costs are achieved when $c^{\text{rel}} = 2$. This makes sense intuitively since assets are relatively expensive, so it is likely that it is more cost effective to invest in spare parts than in spare assets. However, even when $c^{\text{rel}} = 2$, the maximum additional costs are 24%. This clearly shows that joint optimization of spare asset levels and spare LRU levels is necessary in practice, since it is never certain upfront that it is safe to use the lower bound on the asset stock levels, i.e., the lower bound algorithm.

6.4 Computational effort

In this section, we answer Question 4. Table 6 shows the computation times for three algorithms: our algorithm that uses the bound based on Proposition 3, our algorithm that does not use that bound, and the algorithm by De Smidt-Destombes et al. (2011). Since the algorithm of De Smidt-Destombes et al. requires more computation time than our algorithms, we have only run the problem instances of up to 256 LRUs using their algorithm. Note that by construction, all three algorithms find identical solutions.

Given the number of convolutions that both algorithms perform for each LRU in each iteration of the marginal analysis approach, we would expect that, when not using the bound based on Proposition 3, the algorithm of De Smidt-Destombes et al. would require for 16, 64, and 256 LRUs 4, 10.67 and 32 times as much computation time, respectively, being $\frac{|I|}{\log_2 |I|}$. We find that the relative performance of our algorithm is about 70% of what we expected, i.e., their algorithm requires 2.7, 7.6 and 23.5 times as much computation time on average, respectively. This is probably due to our algorithm requiring more storage and overhead. We further see that using the bound that is based on Proposition 3 saves a considerable amount of computation time, with the savings increasing with an increasing problem size: For the instances with 256 LRUs, our algorithm that does not use the bound requires 2.2 times as much computation time on average as our algorithm that does use the bound.

Table 7 shows the computation times on Set 2 for our algorithm that uses the bound. The computation times increase if the number of LRUs or the maximum resupply lead time increases. In both cases, this is due to the fact that more parts need to be stocked, so more iterations of

	# LRUs: $\|I\|$				Rel. costs of an asset: c^{rel}			Max. res. lead t.: $T^{\max}$	
	16	64	256	1,024	0.5	1	2	0.01	0.1
Computation time (seconds)	0.4	3.6	22.6	229.9	95.4	61.5	35.5	5.2	123.1
# Spare asset levels enumerated	4.0	2.9	2.1	1.8	4.2	2.4	1.5	1.6	3.8
Maximum	12	8	5	4	12	7	4	4	12

	Max. ass. time: $\mu^{\max}$		Ave. costs of an LRU: c^{ave}		Target readiness: R^{obj}		
	0.001	0.01	100	1,000	0.9	0.95	0.975
Computation time (seconds)	68.2	60.1	63.0	65.3	61.6	63.9	66.9
# Spare asset levels enumerated	2.8	2.6	2.7	2.7	2.6	2.7	2.8
Maximum	12	11	12	12	10	11	12

Table 7: Set 2: Computation times for all parameters

the greedy algorithm need to be performed. In the former case, an additional reason is that more LRUs need to be checked in each iteration, since Problem (P) is not item-separable. If the relative costs of an asset increase, then the computation times decrease. The key reason for this is that fewer spare asset levels need to be enumerated. The influence of the other parameters on the computation time is limited.

7 Conclusions and recommendations

We have considered the problem of jointly optimizing the number of spare LRUs and spare assets, i.e., the spare parts inventories and fleet size. This is a problem that needs to be solved by companies that use a fleet of assets, e.g., railway operators, shipping companies or defense organizations. We have found that the optimization problem is challenging since it is not item-separable, nor jointly concave. Despite this challenge, we have been able to construct an efficient and effective algorithm. In a numerical experiment, we have shown that our algorithm typically finds solutions that are close to optimal and that our algorithm is relatively fast, both due to the order in which convolutions are performed and due to the bound based on Proposition 3 that is used to avoid performing unnecessary computations.

We have also shown that the asset stock level that our algorithm finds is often close to, or even the same as, the easily computable lower bound in Proposition 1. In fact, in all instances in our numerical experiment with more than 16 LRUs, our algorithm stocks exactly the lower bound, or one more. Since the lower bound is the number of spare assets that are required to achieve the target readiness if there are unlimited spare parts available, our findings imply that spare assets should be acquired to cover the active maintenance time, while they should not be acquired to cover spare parts unavailability. This is the exact opposite of the 'carcass politics' that we mentioned in Section 1: the policy of spending as much of the available budget on buying complete assets, and spending only the remainder on spare parts. Still, setting the

asset stock level to the lower bound and then stocking spare parts can lead to large additional costs. This means that joint optimization is necessary in practice.

It would be interesting to extend our work by modeling the maintenance processes more realistically. We have now assumed that the repair shop that repairs failed parts has ample servers. This may be realistic in many settings, since it can represent lead time agreements with the repair shop, but in other settings it may not be. We have further assumed that assets fail due to a failure in exactly one LRU and that only this LRU is replaced. In practice, often multiple LRUs are replaced at the same time, for example since conditions of components are monitored. It may be interesting to model this. It then becomes relevant to incorporate whether these replacements need to be performed sequentially, or can be performed in parallel.

Another interesting extension would be to consider the optimization of the LRU level itself: In case of a failure in a part of a certain LRU type, it may be possible to either exchange and repair that part, or a module in which the part is contained. This influences the exchange times, the required resources for the exchange, and the types and amounts of spare parts to stock. Some first results on that problem, without considering spare LRUs and spare assets, can be found in Parada Puig and Basten (2015).

Acknowledgements

The authors thank two reviewers, an associate editor, and the department editor, Sila Çetinkaya, for their helpful comments, which enabled us to improve the paper considerably. The authors further thank Bob Huisman of NedTrain for discussions on the real-life situation that led to the model in this paper. The first author gratefully acknowledges the support of the Lloyd's Register Foundation (LRF). LRF helps to protect life and property by supporting engineering-related education, public engagement and the application of research. The second author thanks the Netherlands Foundation for Scientific Research for funding this research.

A Proofs

Proof of Proposition 2

We first derive the difference function for the number of spare LRUs $i \in I$. For notational convenience, let $Z_i = Y_0 + \sum_{k \in I \setminus \{i\}} B_k$. Then:

$$\begin{aligned}
\Delta_i R(\mathbf{S_0}) &= R(\mathbf{S_0} + \mathbf{e}_i) - R(\mathbf{S_0}) \\
&= \left(1 - \mathbb{P}\left\{Z_i + [X_i - S_i - 1]^+ > S_0\right\}\right) - \left(1 - \mathbb{P}\left\{Z_i + [X_i - S_i]^+ > S_0\right\}\right) \\
&= \mathbb{P}\left\{Z_i + [X_i - S_i]^+ > S_0\right\} - \mathbb{P}\left\{Z_i + [X_i - S_i - 1]^+ > S_0\right\} \\
&= \sum_{x=0}^{\infty} \mathbb{P}\left\{Z_i + [X_i - S_i]^+ > S_0 \mid X_i = x\right\} \mathbb{P}\left\{X_i = x\right\} \\
&\quad - \sum_{x=0}^{\infty} \mathbb{P}\left\{Z_i + [X_i - S_i - 1]^+ > S_0 \mid X_i = x\right\} \mathbb{P}\left\{X_i = x\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{x=S_i+1}^{S_i+S_0+1} \mathbb{P}\left\{Z_i + [X_i - S_i]^+ > S_0 \mid X_i = x\right\} \mathbb{P}\left\{X_i = x\right\} \\
&\quad - \sum_{x=S_i+1}^{S_i+S_0+1} \mathbb{P}\left\{Z_i + [X_i - S_i - 1]^+ > S_0 \mid X_i = x\right\} \mathbb{P}\left\{X_i = x\right\} \\
&= \sum_{x=S_i+1}^{S_0+S_i+1} \left[\mathbb{P}\left\{Z_i > S_0 + S_i - x\right\} - \mathbb{P}\left\{Z_i > S_0 + S_i + 1 - x\right\}\right] \mathbb{P}\left\{X_i = x\right\} \\
&= \sum_{x=S_i+1}^{S_0+S_i+1} \mathbb{P}\left\{Z_i = S_0 + S_i + 1 - x\right\} \mathbb{P}\left\{X_i = x\right\} \\
&= \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \mathbb{P}\left\{Z_i = S_0 + 1 - b\right\}.
\end{aligned}
$$

The fifth equation holds because if $X_i < S_i + 1$, then $[X_i - S_i]^+ = [X_i - S_i - 1]^+ = 0$, and if $X_i > S_i + S_0 + 1$, then $\mathbb{P}\left\{Z_i + [X_i - S_i]^+ > S_0\right\} = \mathbb{P}\left\{Z_i + [X_i - S_i - 1]^+ > S_0\right\} = 1$.

We next derive the second order difference function:

$$
\begin{aligned}
\Delta_i^2 R(\mathbf{S_0}) &= \Delta_i R(\mathbf{S_0} + \mathbf{e}_i) - \Delta_i R(\mathbf{S_0}) \\
&= \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + 1 + b\right\} \mathbb{P}\left\{Y_0 + \sum_{k \in I \setminus \{i\}} B_k = S_0 + 1 - b\right\} \\
&\quad - \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \mathbb{P}\left\{Y_0 + \sum_{k \in I \setminus \{i\}} B_k = S_0 + 1 - b\right\} \\
&= \sum_{b=1}^{S_0+1} \left[\mathbb{P}\left\{X_i = S_i + 1 + b\right\} - \mathbb{P}\left\{X_i = S_i + b\right\}\right] \mathbb{P}\left\{Y_0 + \sum_{k \in I \setminus \{i\}} B_k = S_0 + 1 - b\right\}.
\end{aligned}
$$

By Lemma 1, X_i is a Poisson distributed random variable with mean $\lambda_i T_i$ so that we may express its probability mass function recursively as $\mathbb{P}\left\{X_i = k\right\} = \frac{\lambda_i T_i}{k} \mathbb{P}\left\{X_i = k - 1\right\}$, for $k > 0$.

Now, for part (i): If $S_0 + S_i < \lceil \lambda_i T_i \rceil - 2$, then $\mathbb{P}\left\{X_i = S_i + 1 + b\right\} > \mathbb{P}\left\{X_i = S_i + b\right\}$ for $b \in \{1, \ldots, S_0 + 1\}$, so that $\Delta_i^2 R(\mathbf{S_0}) > 0$.

For part (ii): If $S_i \geq \lceil \lambda_i T_i \rceil - 2$, then $\mathbb{P}\left\{X_i = S_i + 1 + b\right\} \leq \mathbb{P}\left\{X_i = S_i + b\right\}$ for $b \in \{1, \ldots, S_0 + 1\}$, so that $\Delta_i^2 R(\mathbf{S_0}) \leq 0$.

Proof of Proposition 3

For notational convenience, let $Z_{ij} = Y_0 + \sum_{k \in I \setminus \{i,j\}} B_k$. Then:

$$
\begin{aligned}
&\Delta_i R(\mathbf{S_0} + \mathbf{e}_j) - \Delta_i R(\mathbf{S_0}) \\
&= \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\left[\mathbb{P}\left\{Z_{ij} + [X_j - S_j - 1]^+ = S_0 + 1 - b\right\} - \mathbb{P}\left\{Z_{ij} + [X_j - S_j]^+ = S_0 + 1 - b\right\}\right] \\
&= \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \sum_{z=0}^{S_0+1-b} \mathbb{P}\left\{Z_{ij} = z\right\} \\
&\quad \left[\mathbb{P}\left\{[X_j - S_j - 1]^+ = S_0 + 1 - b - z\right\} - \mathbb{P}\left\{[X_j - S_j]^+ = S_0 + 1 - b - z\right\}\right] \\
&= \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \sum_{z=0}^{S_0-b} \mathbb{P}\left\{Z_{ij} = z\right\} \\
&\quad \left[\mathbb{P}\left\{[X_j - S_j - 1]^+ = S_0 + 1 - b - z\right\} - \mathbb{P}\left\{[X_j - S_j]^+ = S_0 + 1 - b - z\right\}\right] \\
&\quad + \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \mathbb{P}\left\{Z_{ij} = S_0 + 1 - b\right\} \\
&\qquad \left[\mathbb{P}\left\{[X_j - S_j - 1]^+ = 0\right\} - \mathbb{P}\left\{[X_j - S_j]^+ = 0\right\}\right] \\
&= \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \sum_{z=0}^{S_0-b} \mathbb{P}\left\{Z_{ij} = z\right\} \\
&\quad \left[\mathbb{P}\left\{X_j = S_0 + S_j + 2 - b - z\right\} - \mathbb{P}\left\{X_j = S_0 + S_j + 1 - b - z\right\}\right] \\
&\quad + \mathbb{P}\left\{X_j = S_j + 1\right\} \sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \mathbb{P}\left\{Z_{ij} = S_0 + 1 - b\right\}. \qquad (2)
\end{aligned}
$$

After the third equation, the case that $z = S_0 + 1 - b$ is considered separately. Furthermore, we define $\sum_{x=b}^{c} a_x = 0$ if $c < b$.

We now require two results, which we prove below, that we combine to prove Proposition 3. The first result is that the first of the two terms in Equation (2) is negative, while the second result is that the second term in that equation is smaller than $\mathbb{P}\left\{X_j = S_j + 1\right\} \mathbb{P}\left\{X_i = S_i + 1\right\}$. The summation of the two terms is then also smaller than $\mathbb{P}\left\{X_j = S_j + 1\right\} \mathbb{P}\left\{X_i = S_i + 1\right\}$.

1. X_j in $\mathbb{P}\left\{X_j = S_0 + S_j + 2 - b - z\right\}$ and $\mathbb{P}\left\{X_j = S_0 + S_j + 1 - b - z\right\}$ ranges from $S_j + 1$ and S_j, to $S_j + S_0 + 1$ and $S_j + S_0$, respectively. Since $S_j \geq \lceil \lambda_j T_j \rceil - 2$, the first term in Equation (2) must be negative (due to the properties of the Poisson distribution discussed in the proof of Proposition 2).

2. Since $S_i \geq \lceil \lambda_i T_i \rceil - 2$, it holds that:

$$
\begin{aligned}
&\sum_{b=1}^{S_0+1} \mathbb{P}\left\{X_i = S_i + b\right\} \mathbb{P}\left\{ Y_0 + \sum_{k \in I \setminus \{i,j\}} B_k = S_0 + 1 - b \right\} \\
&< \mathbb{P}\left\{X_i = S_i + 1\right\} \sum_{b=1}^{S_0+1} \mathbb{P}\left\{ Y_0 + \sum_{k \in I \setminus \{i,j\}} B_k = S_0 + 1 - b \right\} \\
&< \mathbb{P}\left\{X_i = S_i + 1\right\}.
\end{aligned}
$$

As a result, the second term in Equation (2) is smaller than $\mathbb{P}\left\{X_j = S_j + 1\right\} \mathbb{P}\left\{X_i = S_i + 1\right\}$.

References

Basten, R. J. I. and Van Houtum, G. J. (2013). Near-optimal heuristics to set base stock levels in a two-echelon distribution network. *International Journal of Production Economics*, 143(2):546–552.

Basten, R. J. I. and Van Houtum, G. J. (2014). System-oriented inventory models for spare parts. *Surveys in Operations Research and Management Science*, 19(1):34–55.

Clark, A. J. and Scarf, H. (1960). Optimal policies for a multi-echelon inventory problem. *Management Science*, 6(4):475–490.

Costantino, F., Di Gravio, G., and Tronci, M. (2013). Multi-echlon, multi-indenture spare parts inventory control subject to system availability and budget constraints. *Reliability Engineering & System Safety*, 119:95–101.

Feeney, G. J. and Sherbrooke, C. C. (1966). The $(s-1, s)$ inventory policy under compound poisson demand. *Management Science*, 12(5):391–411.

Graves, S. C. (1985). A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Science*, 31(10):1247–1256.

Guide Jr., V. D. R. and Srivastava, R. (1997). Repairable inventory theory: Models and applications. *European Journal of Operational Research*, 102:1–20.

Hoff, A., Andersson, H., Christiansen, M., Hasle, G., and Løkketangen, A. (2010). Industrial aspects and literature survey: Fleet composition and routing. *Computers & Operations Research*, 37:2041–2061.

Jin, T. and Wang, P. (2012). Planning performance based contracting considering reliability and uncertain system usage. *Journal of the Operational Research Society*, 63:1467–1478.

Kennedy, W., Wayne Patterson, J., and Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of Production Economics*, 76:201–215.

Muckstadt, J. A. (2005). *Analysis and Algorithms for Service Parts Supply Chains*. Springer, New York (NY).

Palm, C. (1938). Analysis of the erlang traffic formulae for busy-signal arrangements. *Ericsson Technics*, 4:39–58.

Pantuso, G., Fagerholt, K., and Hvattum, L. M. (2014). A survey on maritime fleet size and mix problems. *European Journal of Operational Research*, 235:341–349.

Parada Puig, J. E. and Basten, R. J. I. (2015). Defining line replaceable units. *European Journal of Operational Research*, 247(1):310–320.

Rustenburg, W. D. (2000). *A System Approach to Budget-Constrained Spare Parts Management.* PhD thesis, BETA research school, Eindhoven (The Netherlands).

Safaei, N., Banjevic, D., and Jardine, A. K. S. (2011). Workforce constrained maintenance scheduling for military aircrapft fleet: a case study. *Annals of Operations Research*, 186:295–316.

Salman, S., Cassady, C. R., Pohl, E. A., and Ormon, S. W. (2007). Evaluating the impact of cannibalization on fleet performance. *Quality and Reliability Engineering International*, 23:445–457.

Sherbrooke, C. C. (1968). METRIC: A multi-echelon technique for recoverable item control. *Operations Research*, 16(1):122–141.

Sherbrooke, C. C. (1971). An evaluator for the number of operationally ready aircraft in a multi-level supply system. *Operations Research*, 19(3):618–635.

Sherbrooke, C. C. (2004). *Optimal inventory modelling of systems. Multi-echelon techniques.* Kluwer, Dordrecht (The Netherlands), second edition.

Tjepkema, A. C. (2010). Plannen op cehesie en voorzettingsvermogen in plaats van op deelbelangen. *Carré*, 11/12:30–38. in Dutch.

De Smidt-Destombes, K. S., Van der Heijden, M. C., and van Harten, A. (2004). On the availability of a k-out-of-n system given limited spares and repair capacity under a condition based maintenance strategy. *Reliability Engineering and System Safety*, 83(1):287–300.

De Smidt-Destombes, K. S., van Elst, N. P., Barros, A. I., Mulder, H., and Hontelez, J. A. M. (2011). A spare parts model with cold-standby redundancy on system level. *Computers & Operations Research*, 38(7):985–991.

Van Houtum, G. J. (2006). Multi-echelon production/inventory systems: Optimal policies, heuristics, and algorithms. In Johnson, M. P., Norman, B., and Secomandi, N., editors, *Tutorials in operations research: models, methods, and applications for innovative decision making*, chapter 8, pages 163–199. INFORMS, Hanover (MD).