Spare parts management for technical systems : resupply of spare parts under limited budgets

IIE Transactions (2000) **32**, 1013–1026

Spare parts management for technical systems: resupply of spare parts under limited budgets

W.D. RUSTENBURG[1,2,*], G.J. VAN HOUTUM[2] and W.H.M. ZIJM[2]

[1]*Royal Netherlands Navy, Naval Maintenance Company, PO Box 10 000, 1780 CA Den Helder, The Netherlands*
E-mail: w.d.rustenburg@capitolonline.nl
[2]*Faculty of Technology Management, Eindhoven University of Technology, PO Box 513, 5600 MB, Eindhoven, The Netherlands*

Received July 1999 and accepted January 2000

In this paper, we study the operational availability of a complex technical system consisting of several components. The components are subject to breakdowns, and hence for each component a limited number of spare parts are held in stock. If a system's component fails and it can not be replaced immediately, due to a lack of spares, the system becomes unavailable until a new component is installed. Failed components are disposed of and hence, to keep the spare parts stock at an appropriate level, new components have to be purchased. We assume that only a limited annual budget is available for procurement, while any further procurement requires a considerable lead time. We investigate at an aggregate level what budgets are needed to attain a target availability level for the system. In addition, we develop various operational strategies for spending the annual budget during each year. Numerical results indicate that the so-called Balance Focussed strategy provides the best results in terms of system availability as a function of time.

1. Introduction

The reliability and operational availability of complex technical systems is of crucial importance in many modern manufacturing and service organizations. When such a system fails, due to the failure of one of its components, two options are possible, in principle. The first one is an immediate repair of the failed component. The second one is to replace it by a ready-for-use spare component and to repair the failed item at a more convenient point in time. Due to technological complexity and the need for special equipment and specialized knowledge, the second option is in most cases the more appropriate one. The natural question to be answered then is: how many spare components should be kept in stock to realize a pre-defined target operational availability level for the system? Or, under a given budget constraint for the total investment in spare parts: which components should be held in stock in order to maximize the long term average availability? These questions are already challenging in the rather simple situation of one technical system that consists of components at only one so-called indenture level. A more complicated situation occurs when the bill of materials of the system has several levels and when spare items of components at different levels may be held in stock; this is called the *multi-indenture* case. In addition, the same system may be operational at other locations, in which case repair facilities and spare parts may be required at both the local level and a central level; this is called the *two-echelon* case.

Appropriate models to answer the questions stated above have been developed under the names of METRIC (Multi-Echelon Technique for Recoverable Item Control) and VARI-METRIC models. The research for these models was initiated by Sherbrooke (1968), who considered a two-echelon, single-indenture case. After that, relevant extensions have been made by Muckstadt (1973), Slay (1984), Graves (1985) and Sherbrooke (1986, 1992). These models are formulated in such a way that the maximization of the system's availability, as a function of the target inventory positions for the components, is equivalent to the minimization of the sum of the expected numbers of backorders (for all components and at all downstream locations). The main assumption being made in these models concerns the availability of an unlimited repair capacity, hence failed items do not have to compete for repair capacity. Next, a greedy approach can be used to determine the target inventory positions under which the total expected number of backorders is minimized, subject to a limited budget available for the total number of spare parts to be purchased. In the single-echelon, single-indenture case, the expected number of backorders can be shown to be convex, as a function of these target

*Corresponding author

inventory levels and assuming unlimited repair capacity; as a consequence, the greedy approach can be shown to yield an optimal solution. In the more general cases, optimality can no longer be guaranteed, but the greedy approach still leads to very good results. Thus, METRIC-type models are very suitable to determine how to invest a *one-time given budget* in order to maximize the availability of a system.

A serious drawback of the METRIC-type models developed so far is the fact that they are static in nature, i.e., they assume a stationary stochastic failure behavior, no condemnation (items can always be repaired), and an unchanging system structure. Unfortunately, such assumptions are seldom realistic. First, due to technological progress and the implementation of modifications as a result of earlier maintenance events, an organization will usually attempt to reduce the failure rates. Further, condemnation does occur and new items have to be purchased (and again the failure rates may change). The latter is called *resupply*, as opposed to the *initial supply* discussed above. However, due to political decisions in an organization, the *annual budgets* available for resupply may not be sufficient to replace all condemned items, or, equivalently, to keep up the target availability level as determined initially. This causes the need to rethink the availability level and the desired levels for the inventory positions. Also, being confronted with sometimes even declining budgets for resupply, a company may wonder what will be the best strategy to spend the available budget during a certain year, in order to attain a maximum expected average availability over that year.

This paper discusses resupply decisions under budget constraints, both at a strategic level (what budgets are needed) and at an operational level (how to spend an available budget). It is motivated by problems encountered during the course of a much broader study carried out at a maintenance facility (a dockyard) of the Royal Netherlands Navy (RNN). This maintenance facility is responsible for all overhaul and repair activities of the naval ships (frigates, submarines, minesweepers). In addition, it is responsible for the regular resupply of ships with a sufficient amount of spare parts of both repairable and consumable items. Furthermore, additional spare parts are kept in stock at a central location (close to the dockyard). Since condemnation of repairables regularly occurs, and consumable items by definition are lost after usage and hence have to be procured regularly, the dockyard is supplied with an annual budget, the so-called *resupply budget*, to procure new parts. With that budget, the RNN intends to maintain the inventory positions of all items as close as possible to the inventory positions that were realized during the *initial supply*. The total resupply budget currently varies between 60 and 80 million Netherlands Guilders (i.e., 30 to 40 million US Dollars). It is split into six partial budgets, allocated to certain groups of technical systems. A transfer of money from one group to another one is not possible. Hence, the resupply problem can be decomposed into six independent subproblems.

As indicated above, the decisions on the annual resupply budgets for each of the technical systems groups are often subject to political considerations. Therefore, the total resupply budget may vary over the years. Furthermore, the stochastic parameters of the systems (and thus the demand rates for both consumable and repairable items) may change. Hence the question arises of how a limited resupply budget should be spent over the year in order to achieve a maximum operational availability. A more strategic question is what a sufficient annual resupply budget is, given certain demand characteristics.

This paper aims at answering both questions *for a relative simple situation*. We study the availability of a *single-indenture system* consisting of a number of different components, which, when failed, are replaced by spare items stocked at a *single-location*. To strengthen the role of the limited annual resupply budget, and to simplify the analysis, we assume that all failed items are condemned and hence have to be replaced by new ones. Hence, in this sense, all components can be seen as consumable items.

In the single-indenture, single-location situation, the initial supply problem in its heart reduces to a standard inventory problem, namely the *single-period*, *constrained*, *multi-item inventory problem* (see Silver *et al.* (1998), Section 10.3, and the references therein). Hence, the resupply problem as considered in this paper can be seen as an extension of this standard problem to *an infinite horizon case*. Apart from multiple periods (which seems to make the problem much harder already; (Silver *et al.*, 1998), we also have the property that the annual budget is gradually spent during each period instead of solely at the beginning. Up to now, we have not found any references in which a similar extension has been studied.

The organization and contribution of this paper is as follows. First of all, we outline the rather simple, but characteristic model of the resupply problem in Section 2. This description includes a brief review of the main results for the initial supply problem (cf. Sherbrooke, 1992), which is needed as input for both the formulation the solution of the resupply problem. After that, two topics are studied. In Section 3, we present an aggregate analysis to determine the appropriate long term annual budget size needed for resupply. Next, in Section 4, we develop sensible operational investment strategies for given annual budgets. In particular, we formulate three different strategies (including the one currently used at the RNN), and compare these strategies with respect to their impact on both the average and end-of-year availability. One example that is typical for the technical systems at the RNN, will be used to demonstrate the application of the methods and strategies presented in Sections 2–4. Finally, we end with conclusions and a preview on further research in Section 5.

2. Model description and preliminaries

This section is divided into four parts. First, in Section 2.1, we describe the assumptions with respect to the technical system itself. Next, in Section 2.2, we discuss the initial supply problem and its solution. Subsequently, in Section 2.3, we describe the resupply problem. This part includes the assumptions that are made with respect to the financial restrictions in the exploitation period of the technical system and the definition of the main performance measure. Finally, in Section 2.4, preliminary results for the main performance measure are presented.

2.1. *The technical system*

We consider a (single-indenture) technical system that in essence consists of I different products, that are numbered from $1, \ldots, I$. For each product, one or more items may occur in the configuration of the system. As for all technical systems, we distinguish an *initial phase*, in which the system is procured, and an *exploitation period*, in which the system is operational and performs the task for which it was procured.

In the exploitation period, for each of the I products, items may fail. Usually, only one item fails at a time. A failed item is instantaneously replaced by an item from the spare parts stock, if available. Otherwise, the failed item is replaced as soon as possible, i.e., after a new item has been procured or, if there is at least one item in the procurement pipeline, as soon as the first item in the procurement pipeline has arrived. Between the actual failure and the installation of a new component, the system can not operate and hence is said to be down (note that the downtime can be zero if spare items are available, due to the assumption of instantaneous replacement). In reality, the failure of an item often means that while the item is not in optimal condition anymore, the system can still operate at a lower mode, at least for a short period of time. So, even when the system is (said to be) down, failures continue to occur for all items apart from the one(s) that have failed already. If, in addition, we assume that the times between failures are exponentially distributed, then it follows that for each product $i = 1, \ldots, I$, failures of (all items of) product I occur according to a Poisson process with a constant rate, say m_i failures per year.

To avoid the system being down for excessively long periods after failures of its constituting items, new items may be procured in advance, i.e., new items may be held in stock as spare items. These spare items can be procured both in the initial phase and during the exploitation period. The prices are assumed to be the same in both phases. The price of product i is given by c_i. Items that are procured in the initial phase are available at the beginning of the exploitation period. Items ordered in the exploitation period arrive after a deterministic procurement lead time, that is assumed to be the same for all products. This time is equal to T years. (In the corresponding METRIC-type model, the procurement or repair lead times are allowed to be different for different components and they may be randomly distributed. Here, identical and deterministic lead times are assumed in order to facilitate the analysis of the resupply problem). Further, the fixed ordering costs are assumed to be zero (i.e., negligibly small in comparison to the prices c_i of the products), which justifies a one-for-one replenishment strategy for all products.

Finally, for the sake of clarity, we say that for each product i there is a stockpoint for spares of this product (even when the physical stock of this stockpoint is always zero). Failures of items of this product constitute demands for this stockpoint, and demands that can not be fulfilled immediately are backordered. New items that have been procured are always delivered at this stockpoint (and, in the case of a positive backlog, they will immediately be shipped to replace failed items).

2.2. *The initial supply problem*

We can now formulate the *initial supply problem*. During the initial phase an *initial supply budget* C may be spent on spare items. Let S_i denote the number of spare items of product i to be procured. The total package of spare items is denoted by the vector $\mathbf{S} = (S_1, \ldots, S_I)$. Here, all choices with $S_i \in \mathbb{N}_0$ ($\mathbb{N}_0 := \{0, 1, \ldots\}$) for all i and $\sum_{i=1}^{I} c_i S_i \leq C$ are possible. Obviously, the S_i that are chosen constitute physical stock levels of spares at the beginning of the exploitation period, as well as the inventory positions of the spares, since initially there are no backlogs and no items in the procurement pipeline. The question is how to choose the S_i values. In the initial phase, this is solved as follows. It is *pretended* that in the exploitation period for each failed item one immediately procures a new item to replenish the stock (here, the word pretended has been emphasized, since this intention can not be fulfilled in case only a limited budget is available for the resupply; see Section 2.3). *Under this assumption, the inventory positions always equal the same* (*basestock*) *levels* S_i. The goal is to select the S_i values, given a limited initial supply budget, such that the resulting long run or steady-state fraction of the up-time of the system is maximized. This fraction is called the *availability* and is denoted by $A(\mathbf{S})$.

Below, we briefly describe the analysis and main results for the initial supply problem. The analysis slightly deviates from the one presented by Sherbrooke (1992); for our case with only one instead of multiple technical systems, Sherbrooke's results can be strengthened by using backorder probabilities instead of expected backorders (Rustenburg *et al.*, 1998).

The first part of the analysis consists of the derivation of an expression for the availability $A(\mathbf{S})$. The exploitation period starts at time $t = 0$, and we pretend that it lasts infinitely long. Consider an arbitrary time $t \geq T$ in the exploitation period. The system is down at this time if and only if there is at least one item in its configuration that has failed and has not yet been replaced because there is no spare item available, i.e., if and only if there is a backorder for at least one of the products. Since at time $t - T$ the inventory position for product i was equal to S_i, a backorder at time t occurs if and only if in the time interval $(t - T, t]$ the number of failures for product i is larger than S_i. This number of failures is Poisson distributed with mean $m_i T$, and hence the probability of any backorder for product i at time t is equal to

$$PBO_i(S_i) = \sum_{k=S_i+1}^{\infty} \frac{(m_i T)^k}{k!} \mathrm{e}^{-m_i T} = 1 - \sum_{k=0}^{S_i} \frac{(m_i T)^k}{k!} \mathrm{e}^{-m_i T}.$$

The probability that there is no backorder at time t equals $1 - PBO_i(S_i)$ for product i. Hence the probability that the system is up at time t equals $\prod_{i=1}^{I}(1 - PBO_i(S_i))$. Since the same reasoning can be applied for any $t \geq T$, the availability $A(\mathbf{S})$ also satisfies:

$$A(\mathbf{S}) = A(S_1, \ldots, S_I) = \prod_{i=1}^{I}(1 - PBO_i(S_i)). \quad (1)$$

The second part of the analysis starts with formulating the goal in terms of a Non-Linear Integer Programming (NLIP) problem:

$$\text{Max } A(S_1, \ldots, S_I), \quad (2)$$

subject to:

$$\sum_{i=1}^{I} c_i S_i \leq C, \quad S_i \in \mathbb{N}_0 \text{ for all } i = 1, \ldots, I.$$

Clearly, maximizing $A(S_1, \ldots, S_I)$ is the same as maximizing its logarithm. Using (1) and approximating $\log(1 - z)$ by the first term of its Taylor expansion yields:

$$\log A(S_1, \ldots, S_I) \approx -\sum_{i=1}^{I} PBO_i(S_i).$$

This approximation will be accurate if all $PBO_i(S_i)$ are sufficiently small, which will be true for all relevant choices for the S_i. Therefore, instead of solving (2), it makes sense to solve the following NLIP problem:

$$\text{Min} \sum_{i=1}^{I} PBO_i(S_i), \quad (3)$$

subject to:

$$\sum_{i=1}^{I} c_i S_i \leq C, \quad S_i \in \mathbb{N}_0 \text{ for all } i = 1, \ldots, I.$$

(This problem is a variant of the single-period, constrained, multi-item inventory problem as described in Section 10.3 of Silver *et al.* (1998) and it is solved in a similar way).

The objective function of problem (3) is equal to the sum of I independent terms where the i-th term solely depends on S_i. For each i, the i-th term of $PBO_i(S_i)$ is strictly decreasing for all $S_i \geq 0$. In addition, one may show that $PBO_i(S_i)$ is convex for $S_i \geq \max\{\lceil m_i T - 2 \rceil, 0\}$, i.e., for all $S_i \geq 0$ if $m_i \leq 2/T$, and for all values excluding the smallest values $S_i < \lceil m_i T - 2 \rceil$ if $m_i > 2/T$.

The next step is to exclude the latter values $S_i < \lceil m_i T - 2 \rceil$ from the solution space. By this step, we obtain the following NLIP problem:

$$\text{Min} \sum_{i=1}^{I} PBO_i(S_i), \quad (4)$$

subject to:

$$\sum_{i=1}^{I} c_i S_i \leq C,$$

$$S_i \geq \max\{\lceil m_i T - 2 \rceil, 0\} \text{ and integer for all } i = 1, \ldots, I.$$

The objective function of this problem consists of independent terms that are strictly decreasing and convex for all feasible values. This allows a *greedy approach* to solve this NLIP problem. The justification for the reduction of the solution space is as follows. For the products i with $m_i > 2/T$, a value $S_i < \lceil m_i T - 2 \rceil$ will lead to a large value for $PBO_i(S_i)$ (notice that $m_i T$ is the mean of the Poisson distribution from which $PBO_i(S_i)$ is obtained and that the corresponding Poisson probabilities have their maximum at $\lceil m_i T \rceil$). Hence, an unattractive value for the availability $A(\mathbf{S})$ is obtained in that case (see (1)). So, for reasonable values of the budget C, i.e., values for which a sufficiently high availability $A(\mathbf{S})$ can be reached, the optimal solution(s) of (4) may be expected to be the same as for (3).

By applying a greedy approach to the NLIP problem stated in (4), we obtain a series of solutions $\mathbf{S}^p = (S_1^p, \ldots, S_I^p)$, $p \in \mathbb{N}$, such that for each p the vector $\mathbf{S}^p$ is the optimal solution of (4) with $C = C_p = \sum_{i=1}^{I} c_i S_i^p$. The solutions $\mathbf{S}^p$ are generated as follows. Define $\mathbf{S}^1 = (S_1^1, \ldots, S_I^1)$ by $S_i^1 = \max\{\lceil m_i T - 2 \rceil, 0\}$, $i = 1, \ldots, I$. Next, for each $p \in \mathbb{N}$, $\mathbf{S}^{p+1}$ is obtained from $\mathbf{S}^p$ by first computing

$$\Delta_i = \frac{PBO_i(S_i^p) - PBO_i(S_i^p + 1)}{c_i}, \quad \text{for } i = 1, \ldots, I,$$

and then defining $S_j^{p+1} := S_j^p + 1$ for that index j that maximizes Δ_j, while $S_i^{p+1} := S_i^p$ for all $i \neq j$. Note that, for any i, Δ_i equals the decrease in the sum of the backorder probabilities per extra invested NLG if S_i^p would be increased to $S_i^p + 1$. Hence, in each step the procedure calculates the optimal return, in terms of backorder reduction, per unit of money invested. The procedure is continued until the available budget C is (almost) consumed. The last solution $\mathbf{S}^P$, which is denoted by the index P, can be shown (using convexity properties) to be the optimal solution for the NLIP stated in (4) if we neglect a possible difference between C_P and C. This difference will be negligibly small for all relevant problem sizes.

The vectors $\mathbf{S}^p$, $p = 1, \dots, P$, are also solutions for the original NLIP problem stated in (2). The corresponding pairs (C_p, A_p) of investment C_p and availability $A_p = A(\mathbf{S}^p)$ can not be guaranteed to be optimal anymore, but for large indices p corresponding to high values for C_p and A_p, the solutions (C_p, A_p) may be expected to be close-to-optimal (since the transformations from NLIP problem (2) to (3) and from NLIP problem (3) to (4) will not affect the optimal solution(s) for large budgets C). The solutions $\mathbf{S}^p$ will play an important role in the analysis of the resupply problem. Notice that the corresponding investments and availabilities are increasing, i.e., $C_{p+1} > C_p$ and $A_{p+1} > A_p$ for all $p = 1, \dots, P-1$ (since $\mathbf{S}^{p+1} \geq \mathbf{S}^p$ and $S_i^{p+1} > S_i^p$ for at least one i). In the description above, the generation of the $\mathbf{S}^p$ was stopped when the investment reached the given initial supply budget C. Obviously, one may also decide to stop when a given target availability has been reached.

Example 1: A fire extinguishing system

As an illustrative example, we consider a fire extinguishing system that is in use at the RNN. This system consists of three pump units, where each pump unit consists of seven components. Although these components have the same names for all three pump units, they do differ from each other. Hence, the complete system consists of $I = 21$ different products, all of which are subject to failures. The failure rates (in failures per year) and prices (in NLG, which stands for Netherlands Guilders) for these products are given in Table 1. The procurement lead time is equal to $T = 0.4$ years for all products.

We have applied the greedy approach as described above to this problem until a solution with an availability of at least 97.50% was reached. In that way, we have obtained $P = 127$ solution vectors $\mathbf{S}^p$. The first solution was equal to

$$\mathbf{S}^1 = (0,0,1,2,1,0,2,\ 0,0,1,2,0,0,3,\ 0,0,0,1,1,0,2),$$

where the first seven positions correspond to the components of pump unit 1 (following the order of Table 1), the next seven positions to the components of pump unit 2, and the last seven positions to the components of pump unit 3. The corresponding investment and availability are equal to $C_1 = 7270$ NLG and $A_1 = 0.00\%$. For the last solution with index $P = 127$, we found

$$\mathbf{S}^P = (2,2,9,11,8,7,11,\ 2,1,8,10,7,7,12,\ 3,2,7,9,9,6,10),$$

with $C_P = 87\,720$ NLG and $A_P = 97.54\%$. The pairs (C_p, A_p), $p = 1, \dots, P$, that constitute a so-called investment versus availability curve, are graphically depicted in Fig. 1.

2.3. *The resupply problem*

Once the initial supply problem has been solved and its solution has been implemented, the exploitation period starts and therefore also the resupply problem. For the resupply problem we will make use of the series of solutions $\mathbf{S}^p$ that have been generated for the initial supply problem. The resupply problem is described as follows.

First of all, each time instant in the exploitation period is denoted by a pair (y, t), where $y \in \mathbb{N}$ denotes the y-th year in the exploitation period and $t \in [0, 1]$ denotes the time instant within the y-th year. Here, $(y, 0)$ corresponds to the beginning of year y, and $(y, 1)$ corresponds to the end of year y and just precedes $(y+1, 0)$. We assume that during the initial supply the solution $\mathbf{S}^P$ has been chosen and implemented. (Alternatively, without complicating the analysis, one may assume that one of the other solutions $\mathbf{S}^p$ has been selected during the initial supply). This means that at time instant $(0, 0)$, the physical stock of spares of product $i = 1, \dots, I$ is equal to S_i^P, while both the backlog and the number of items in the procurement pipeline equals zero. Thus, the inventory position of product i is also equal to S_i^P at time instant $(0, 0)$.

If enough money is always available to purchase a new item for each failed item, then the inventory positions could always be kept equal to the levels S_i^P. However, each year only a limited amount of money is available. We assume that each year y, a budget B is available to procure new items. This budget becomes available at the

Table 1. Input data for Example 1

Name	*Pump unit 1*		*Pump unit 2*		*Pump unit 3*	
	m_i *(Failures/yr)*	c_i *(NLG)*	m_i *(Failures/yr)*	c_i *(NLG)*	m_i *(Failures/yr)*	c_i *(NLG)*
Pump	0.8	2230	0.7	2510	0.9	2060
Elmo	0.4	3770	0.3	3990	0.6	3870
Bearing	6.1	330	5.2	330	4.7	330
Seal	9.2	450	8.7	450	6.8	450
Casing	5.4	480	4.5	480	6.7	480
Rotor	4.2	250	3.7	250	3.2	250
Stator	9.8	450	10.5	450	8.9	450

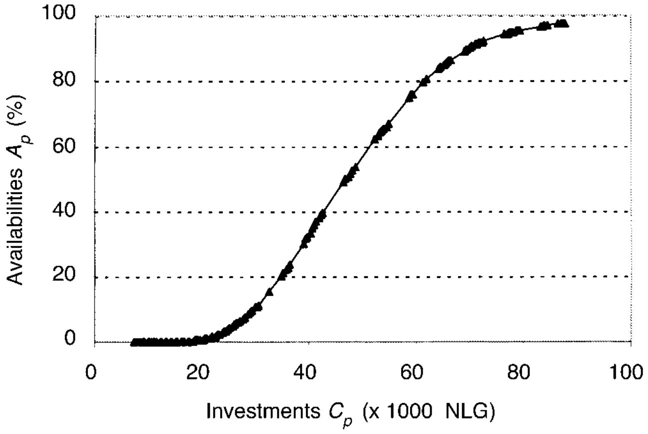

Fig. 1. The investment versus availability curve as constituted by the pairs (C_p, A_p) for Example 1.

beginning of each year y, i.e., at $(y, 0)$, and limits the cumulative value of all items that can be *ordered* during year y. The underlying assumption is that for each new item that is ordered an amount of money has to be reserved from the budget and that money is used to pay for the item when it is delivered. For the sake of clarity, orders can be placed during the whole year, and not just at the beginning or the end of a year. How many items are procured, which ones, and at what time instants during a year, depends on the *operational investment strategy*.

To avoid excessive inventory positions (and waste of money, as often seen in practice at the end of a year when some people order goods just because they want to spend any remaining budget), we assume that it is not allowed to increase the inventory positions to higher levels than the levels S_i^P of the ultimate solution $\mathbf{S}^P$. Notice that we have also assumed that $\mathbf{S}^P$ represents the solution that has been chosen during the initial supply. (But, as stated above, we could assume that for the initial supply another solution $\mathbf{S}^p$ with a smaller index $p < P$ has been taken). Because of the limits S_i^P for the inventory positions, for all strategies the long run fraction of up-time of the system is bounded from above by A_P. Finally, we assume that at the end of each year any money in the budget that has not been spent or reserved for items in the procurement pipeline, is lost. Under sensible strategies, at the end of each year:

(i) either all inventory positions are equal to or increased up to the maximum levels S_i^P and some budget is lost;
(ii) or it is not possible to get all inventory positions equal to their maximum levels but then the whole budget has been spent (apart from a negligibly small amount of money that may remain because one can not procure fractional items).

Let, for $i = 1, \ldots, I$, the generic random variable $\underline{D}_i$ denote the cumulative value of all failed items of product i during a year, and let $\underline{D} = \sum_{i=1}^{I} \underline{D}_i$. Clearly, $\underline{D}_i$ is equal to the product of c_i and a Poisson distributed random variable with mean m_i. Hence, we find that the mean and variance of $\underline{D}$ are equal to

$$E\{\underline{D}\} = \sum_{i=1}^{I} E\{\underline{D}_i\} = \sum_{i=1}^{I} c_i m_i, \tag{5}$$

$$\text{Var}\{\underline{D}\} = \sum_{i=1}^{I} \text{Var}\{\underline{D}_i\} = \sum_{i=1}^{I} c_i^2 m_i. \tag{6}$$

If $B > E\{\underline{D}\}$, then, under any sensible strategy, the inventory positions will always return to their maximum levels eventually (although possibly not every year) and the average amount of money that is lost at the end of the year will be equal to $B - E\{\underline{D}\}$. If $B \leq E\{\underline{D}\}$, then the inventory positions will not return to their maximum levels after a while and the average amount of money that is lost at the end of the year will be equal to zero in the long run. In the latter case, the long run fraction of up-time of the system will be equal to zero for all strategies, but the fraction of up-time during the first years of the exploitation period may still be reasonably good.

As for the initial supply problem, the target for the resupply problem is to maximize the availability of the technical system. However, we are not interested in the long run fraction of up-time of the system, but in the up-times in the first 30 years, say, of the exploitation period. Define $\tilde{A}(y, t)$ as the probability that the system is up at time (y, t) of the exploitation period. $\tilde{A}(y, t)$ is called the *availability probability*, or in short the *availability*. Obviously, the availability $\tilde{A}(y, t)$ depends on the failures that occur for each of the products $i = 1, \ldots, I$ and the investment strategy that is being used. The goal is to find a strategy such that the function $\tilde{A}(y, t)$ is as high as possible over a period of $Y \in \mathbb{N}$ years, starting from the initialization of the system. For the dockyard, for instance, $Y = 30$ years. Hence, instead of comparing different strategies on the basis of one number (e.g., the average availability, we are merely interested in the behavior of the availability $\tilde{A}(y, t)$ over the whole exploitation period up to and including the Y-th year. In the next section, we study the availability probability $\tilde{A}(y, t)$ in more detail.

2.4. *Preliminary results for the availability* $\tilde{\mathbf{A}}(\mathbf{y}, \mathbf{t})$

In this Section, we show that $\tilde{A}(y, t)$ directly depends on the inventory positions of the products $i = 1, \ldots, I$. Then, this result is used to define related performance measures that we will focus on in the remainder of this paper.

In general, the availability $\tilde{A}(y, t)$ is positively affected by the procurement of new items. However, the time $(1, 0)$ is the first one on which items can be ordered and thus only after T years can the first new items arrive. Hence, the availability during the first T years only depends on the physical stocks S_i^P with which we start at time $(1,0)$. To simplify the formulae below, we assume that the procurement lead time $T \le 1$ (year). Then, similar to (1), we find

$$\tilde{A}(1,t) = \prod_{i=1}^{I} \sum_{k=0}^{S_i^P} \frac{(m_i t)^k}{k!} e^{-m_i t}, \quad 0 \le t < T. \tag{7}$$

For all time instants (y,t) after the first T years of the exploitation period, the availability $\tilde{A}(y,t)$ follows directly from the inventory positions of the products $i = 1,\ldots,I$ at the time T years earlier, i.e., at time $(y,t-T)$ if $y \ge 1$ and $T \le t < 1$ and at time $(y-1,t-T+1)$ if $y \ge 2$ and $0 \le t < T$. These inventory positions are affected by the procurements up to this time and these procurements depend on the investment strategy that is being used and the times at which failures have occurred in that period.

By the random variables $\underline{X}_i(y,t)$, we denote the inventory positions of the products i at time (y,t), while the variables $x_i(y,t)$ denote realizations of the $\underline{X}_i(y,t)$. Further, let $\underline{\mathbf{X}}(y,t) := (\underline{X}_1(y,t),\ldots,\underline{X}_I(y,t))$ and $\mathbf{x}(y,t) := (x_1(y,t),\ldots,x_I(y,t))$. Let $y \ge 1$ and $T \le t < 1$. Then, again similar to (1), we obtain the following formula for the availability $\tilde{A}(y,t)$, under the condition that $\underline{\mathbf{X}}(y,t-T) = \mathbf{x}(y,t-T)$:

$$\tilde{A}(y,t|\underline{\mathbf{X}}(y,t-T) = \mathbf{x}(y,t-T)) = A(\mathbf{x}(y,t-T)),$$

where the function $A(\cdot)$ is defined by (cf. (1)):

$$A(\mathbf{x}) = A(x_1,\ldots,x_I) = \prod_{i=1}^{I}(1 - PBO_i(x_i)), \quad \mathbf{x} \in \mathbb{N}_0^I, \tag{8}$$

with

$$PBO_i(x_i) = \sum_{k=x_i+1}^{\infty} \frac{(m_i T)^k}{k!} e^{-m_i T} = 1 - \sum_{k=0}^{x_i} \frac{(m_i T)^k}{k!} e^{-m_i T},$$

and $A(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{Z}^I$ with at least one component $x_i < 0$. Integration over all possible realizations of $\underline{X}_i(y,t)$ yields

$$\begin{aligned}\tilde{A}(y,t) &= \sum_{\mathbf{x}\in\mathbb{Z}^I} A(\mathbf{x})\Pr\{\underline{\mathbf{X}}(y,t-T) = \mathbf{x}\} \\ &= E\{A(\underline{\mathbf{X}}(y,t-T))\}, \quad y \ge 1, \ T \le t < 1.\end{aligned}$$

The expression on the right-hand side is also denoted as $A^{\mathrm{IP}}(y,t-T)$ and is called the *IP-based availability probability* (or *IP-based availability*) at time $(y,t-T)$, where IP stands for Inventory Position. To distinguish the IP-based availability probability from the availability probability $\tilde{A}(y,t)$, the latter from now on is called the *actual availability*. It is important to note the difference between the two notions: $\tilde{A}(y,t)$ denotes the availability of the system at time (y,t), whereas $A^{\mathrm{IP}}(y,t)$ is related to backorder positions at time (y,t) and thereby determines the availability of the system T years later.

Similarly, it follows that:

$$\begin{aligned}\tilde{A}(y,t) &= E\{A(\underline{\mathbf{X}}(y-1,t-T+1))\} \\ &= A^{\mathrm{IP}}(y-1,t-T+1), \quad y \ge 2, \ 0 \le t < T.\end{aligned}$$

This completes the proof of the following lemma.

Lemma 1. *The availability probability* $\tilde{A}(y,t)$ *is equal to*

$$\tilde{A}(y,t) = \begin{cases} A^{\mathrm{IP}}(y,t-T) & \text{for } y \ge 1 \text{ and } T \le t < 1; \\ A^{\mathrm{IP}}(y-1,t-T+1) & \text{for } y \ge 2 \text{ and } 0 \le t < T, \end{cases}$$

where the IP-based availability $A^{\mathrm{IP}}(y,t)$ *is defined by* $A^{\mathrm{IP}}(y,t) := E\{A(\underline{\mathbf{X}}(y,t))\}$ *for all time instants* (y,t), $A(\cdot)$ *is given by* (8), *and* $\underline{\mathbf{X}}(y,t)$ *denotes the inventory positions of the products* $i = 1,\ldots,I$ *at time instant* (y,t).

In words, Lemma 1 states that the behavior of $\tilde{A}(y,t)$ is identical to the behavior of $A^{\mathrm{IP}}(y,t)$ but delayed by T years. This is visualized in Fig. 2, where for Example 1 (with $B = 1.05 \times E\{\underline{D}\} = 53\,235$ NLG) and a given investment strategy, the simulated behavior of both the IP-based availability $A^{\mathrm{IP}}(y,t)$ and the actual availability $\tilde{A}(y,t)$ is displayed over three consecutive years of the exploitation period (years 15, 16 and 17). The strategy that has been used here is a simple one: in each year y the budget B is used to keep the inventory positions at the levels S_i^P as long as possible (this is one of the strategies discussed in Section 4). As we see, under this strategy, the IP-based availability $A^{\mathrm{IP}}(y,t)$ in each year is high in the first part, after which it decreases to much lower values at the end of the year. For the actual availability $\tilde{A}(y,t)$, we observe the same behavior, with a delay of T years. However, notice that for $\tilde{A}(y,t)$ in each year y the lowest values are obtained in the first part up to time instant (y,T) instead of at the end of the year.

From now on, we shall only consider the behavior of the IP-based availability $A^{\mathrm{IP}}(y,t)$. Further, we shall

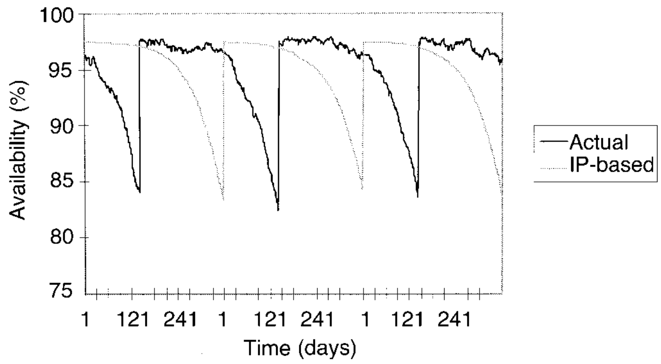

Fig. 2. Realization of the IP-based availability $A^{\mathrm{IP}}(y,t)$ and the actual availability $\tilde{A}(y,t)$ over three consecutive years of the exploitation period under a given investment strategy for Example 1.

mainly focus on the *average IP-based availability per year*, which is defined by

$$A^{\mathrm{IP}}_{\mathrm{avg}}(y) := \int_0^1 A^{\mathrm{IP}}(y,t)\,\mathrm{d}t, \quad y \in \mathbb{N}, \tag{9}$$

and the *end-of-year IP-based availability* for each year, which is defined by

$$A^{\mathrm{IP}}_{\mathrm{eoy}}(y) := A^{\mathrm{IP}}(y,1), \quad y \in \mathbb{N}. \tag{10}$$

When comparing different strategies, high values for $A^{\mathrm{IP}}_{\mathrm{avg}}(y)$ are most important. However, $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$ is also important since in general it represents the IP-based availability at its lowest level. Excessively low values for this measure are undesirable; they represent an inferior system availability.

3. Upper bound for the end-of-year availability

Under sensible investment strategies, at the end of each year either the whole budget has been spent or the inventory positions of all products are equal to their maximum levels S_i^P. Due to this property, the distribution functions of the total invested capital in spare items at the end of the years $y = 1, 2, \ldots$ are the same for all sensible strategies. In this section, we first show that these distributions can be determined by a recursive procedure. Next, we exploit these distributions to obtain upper bounds for the IP-based end-of-year availabilities $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$. These upper bounds can be computed rather easily and quickly. These bounds may be used to obtain a quick first impression on the influence of the extent of the annual budget B on the behavior of the availability.

Let the random variable $\underline{C}(y)$, $y \in \mathbb{N}$, denote the investment in spare items at the end of year y of the exploitation period. Here, backlogs are seen as negative investments, and hence

$$\underline{C}(y) = \sum_{i=1}^{I} c_i \underline{X}_i(y,1), \quad y \in \mathbb{N}.$$

In addition, we define $\underline{C}(0)$ as the investment at the beginning of the exploitation period, i.e., $\underline{C}(0) = \sum_{i=1}^{I} c_i \underline{X}_i(1,0) = C_P$.

The invested capital $\underline{C}(y+1)$ at the end of year $y+1$, $y \in \mathbb{N}_0$, follows from $\underline{C}(y)$ in the following way. If no new items are procured in year $y+1$, then the invested capital at the end-of-year $y+1$ is equal to $\underline{C}(y) - \underline{D}$. However, under each sensible strategy, the budget B will be spent as much as possible, since all money that is not spent in year $y+1$, will be lost. Recall that the inventory positions may not exceed the levels S_i^P, or, equivalently, that the capital investment can not exceed C_P. So, under each sensible strategy, it holds that

$$\underline{C}(y+1) = \min\{\underline{C}(y) - \underline{D} + B, C_P\}, \quad y \in \mathbb{N}_0.$$

This recursive relation may be rewritten as

$$\begin{aligned} C_P - \underline{C}(y+1) &= C_P - \min\{\underline{C}(y) - \underline{D} + B, C_P\} \\ &= -\min\{\underline{C}(y) - \underline{D} + B - C_P, 0\} \\ &= \max\{0, (C_P - \underline{C}(y)) + \underline{D} - B\} \\ &= [(C_P - \underline{C}(y)) + \underline{D} - B]^+, \quad y \in \mathbb{N}_0. \end{aligned} \tag{11}$$

Here $C_P - \underline{C}(y)$ represents the *shortfall* with respect to the maximum capital investment C_P at the end-of-year y.

Let F_y be the distribution function of $C_P - \underline{C}(y)$, $y \in \mathbb{N}_0$. Since $C_P - \underline{C}(0) = 0$, for F_0 the whole probability mass is concentrated in zero, i.e., $F_0(x) = 0$ for all $x < 0$ and $F_0(x) = 1$ for all $x \geq 0$. Next, for each $y \in \mathbb{N}_0$, F_{y+1} follows from F_y as follows. The distribution function of $(C_P - \underline{C}(y)) + \underline{D}$ is given by $F_y * F_D$, where F_D represents the distribution function for $\underline{D}$ and $*$ represents the convolution operator for two distribution functions of two independent random variables. Hence,

$$F_{y+1}(x) = \begin{cases} (F_y * F_D)(x+B) & \text{if } x \geq 0; \\ 0 & \text{if } x < 0. \end{cases} \tag{12}$$

The distribution F_{y+1} is obtained by shifting $F_y * F_D$ to the left over a distance B, while at the same time all probability mass that tries to pass the point $x = 0$ is absorbed at this point. This distribution is also denoted by $(F_y * F_D)^B$, and relation (12) then reads as

$$F_{y+1} = (F_y * F_D)^B, \quad y \in \mathbb{N}_0. \tag{13}$$

In fact, this relation is equivalent to (11).

In principle, each F_y is a discrete distribution. However, for all practical cases, F_D is close to a continuous distribution, and then also each F_y is a continuous distribution, but still with a positive probability mass in zero (i.e., there is a positive probability that there is no shortfall at the end of year y).

We now can formulate the following lemma for the invested capital $\underline{C}(y)$ at the end of year y.

Lemma 2. *Under each sensible investment strategy, for all* $y \in \mathbb{N}_0$, *the probability distribution function of* $\underline{C}(y)$ *satisfies*

$$\Pr\{\underline{C}(y) \leq x\} = \begin{cases} 1 - F_y(C_P - x) & \text{if } x < C_P; \\ 1 & \text{if } x \geq C_P, \end{cases}$$

where the distribution function F_y *is determined recursively via* (13).

The shortfall process $\{C_P - \underline{C}(y)\}_{y \geq 0}$ is identical to both the waiting time process of a $G|D|1$ queue and the shortfall process for the inventories at a periodic-review single-stage capacitated inventory system that operates under a basestock policy (van Houtum and Zijm, 1997). First of all, from this property we learn that the average shortfall is mainly determined by the 'workload' $E\{\underline{D}\}/B$ (which may be ≥ 1) and the coefficient of variation of $\underline{D}$. Second, we can exploit this property to compute the distributions for the shortfall distributions F_y, and thus also for the distributions of the $\underline{C}(y)$. For the shortfalls in the single-stage capacitated inventory problem, both an exact and an approximate procedure has been derived in van Houtum and Zijm (1991,1997). The exact procedure could also be applied here if the distribution F_D is equal to a mixture of one or more Erlang distributions with the same scale parameter. In most practical cases, F_D will have a rather small coefficient of variation and a shape that is close to the shape of an $E_{k-1,k}$ distribution, which is a mixture of an Erlang$(k-1)$ and an Erlang(k) distribution with the same scale parameter. Hence, in that case, an $E_{k-1,k}$ distribution could be fitted on the first two moments of F_D and next an exact procedure could be applied to determine the distributions F_y, $y \in \mathbb{N}$. However, we prefer to use the approximate procedure, since it is known to be very accurate if the shape of F_D is close to the shape of an $E_{k-1,k}$ distribution, and also it is much faster and easier to use.

The approximate procedure is based on a recursive calculation of first and second moments of all distribution functions F_y. The first two moments of F_0 are both equal to zero.

Suppose, we have obtained the first two moments of F_y. From these, the first two moments of $F_y * F_D$ are easily computed. Next, we fit either an $E_{k-1,k}$ or a hyperexponential distribution on these first two moments (depending on whether the corresponding coefficient of variation is smaller than or larger than one), after which the first two moments of the shifted distribution $(F_y * F_D)^B = F_{y+1}$ are easily calculated. Finally, an approximation for the distribution F_{y+1} is obtained by fitting an $E_{k-1,k}$ or hyperexponential distribution on its first two moments. For details on the approximate procedure, the reader is referred to van Houtum and Zijm (1991).

We now can exploit the results stated above to obtain an *upper bound* for the end-of-year availabilities $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$. At the end-of-year y , the invested capital $\underline{C}(y)$ is equal to C_P with a probability $F_y(0)$. In that case, for each product i the inventory position is equal to its maximum S_i^P and the corresponding IP-based availability is equal to A_P. Further, for each $p = 2, \ldots, P$, $\underline{C}(y) \in (C_{p-1}, C_p]$ (exclude C_p for $p = P$) with probability

$$\begin{aligned}\Pr\{C_{p-1} < \underline{C}(y) \le C_p\} &= [1 - F_y(C_P - C_p)] \\ &\quad - [1 - F_y(C_P - C_{p-1})] \\ &= F_y(C_P - C_{p-1}) - F_y(C_P - C_p),\end{aligned}$$

in which case the corresponding IP-based availability lies between A_{p-1} and A_p, and is bounded from above by A_p. Furthermore, $\underline{C}(y) \in [0, C_1]$ with probability $F_y(C_P) - F_y(C_P - C_1)$, while the corresponding IP-based availability is bounded from above by A_1. Finally, when $\underline{C}(y) < 0$, at least one of the products faces a backlog, hence the IP-based availability is equal to zero. All these observations together lead to the upper bound stated in Theorem 1 below.

In fact, the above observations prove the upper bound for all sensible strategies. In Theorem 1, the upper bound is also defined for $y = 0$. We define $A^{\mathrm{IP}}_{\mathrm{eoy}}(0)$ as the availability at the beginning of the exploitation period, i.e., $A^{\mathrm{IP}}_{\mathrm{eoy}}(0) = A_P$ (hence, for $y = 0$, the upper bound is equal to $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$). Once the distribution functions F_y have been computed, the computation of the upper bound for all relevant y is straightforward.

Theorem 1. *For each investment strategy, the IP-based end-of-year availability* $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$, $y \in \mathbb{N}_0$, *satisfies*

$$\begin{aligned}A^{\mathrm{IP}}_{\mathrm{eoy}}(y) \le F_y(0)A_P & \\ &+ \sum_{p=2}^{P} \{F_y(C_P - C_{p-1}) - F_y(C_P - C_p)\}A_p \\ &+ \{F_y(C_P) - F_y(C_P - C_1)\}A_1.\end{aligned}$$

The upper bounds for the $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$ will be accurate for strategies under which the annual budget is invested such that the inventory positions at the end of the year are equal to or close to one of the vectors $\mathbf{S}^p$. In that case, the inventory positions are said to be *balanced*. Obtaining balanced inventory positions at the end of the year requires that a sufficient amount of money is left during the last part of the year to recover from unbalanced situations. However, for the average availabilities $A^{\mathrm{IP}}_{\mathrm{avg}}(y)$, it is better to spend money at the beginning of the year, since that is beneficial for the IP-based availability during the whole year instead of only during the last part of the year. So, for strategies that are good with respect to the average availabilities $A^{\mathrm{IP}}_{\mathrm{avg}}(y)$, the upper bounds for the end-of-year availabilities $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$ are not expected to be accurate (and this is confirmed in Section 4). Nevertheless, these upper bounds still indicate the order of magnitude of the (end-of-year) availabilities that may be expected for a given annual budget, as well as the budget size required to attain an acceptable availability level.

Example 1: A fire extinguishing system (continued)

For Example 1 with varying budgets B, we have depicted the upper bounds for $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$, $y = 0, 1, \ldots, 30$, in Fig. 3. The budget has been taken equal to $B = aE\{\underline{D}\}$, where $E\{\underline{D}\} = 50\,700$ NLG and $a = 0.98, 1.00, 1.02, 1.05, 1.10$. The computations were executed on a Pentium II, 266 MHz. The total computation time required for all five curves of Fig. 3, was equal to about 1 second.

The results show that for $a = 0.98$ and $a = 1.00$, the upper bound for $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$ continuously decreases for increasing values of y. This is in agreement with the properties stated immediately after Equations (5) and (6). For each of the other three values of a, the upper bound for $A^{\mathrm{IP}}_{\mathrm{eoy}}(y)$ decreases to its limit value, which is equal to 97.0% for $a = 1.10$, 95.2% for $a = 1.05$ and 82.9% for

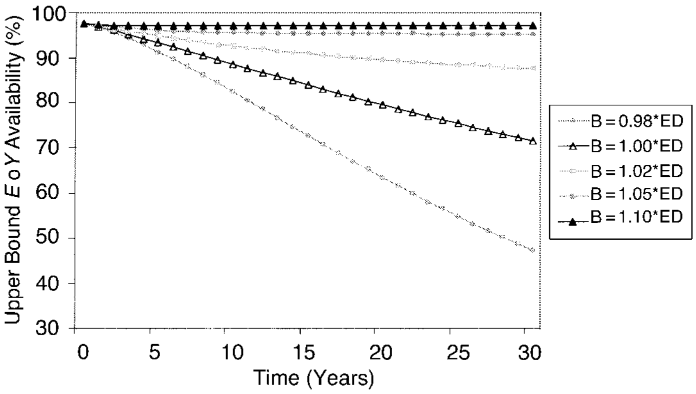

Fig. 3. The upper bound for $A_{\text{eoy}}^{\text{IP}}(y)$, $y = 0, 1, \ldots, 30$, for Example 1 with varying values for the annual budget B.

$a = 1.02$ (for this value of a it takes a long time until $A_{\text{eoy}}^{\text{IP}}(y)$ reaches its limit value; after 30 years, $A_{\text{eoy}}^{\text{IP}}(y)$ is still equal to 87.5%. From the figure, we may conclude that for availabilities of 90% or more a budget of at least $B = aE\{\underline{D}\}$ with $a = 1.02$ is needed. This value for a is rather low, which is due to the fact that for Example 1, $\text{Var}(\underline{D}) = 48\,830\,500$ NLG^2 and the coefficient of variation is equal to $cv\{\underline{D}\} = 0.138$. If this coefficient of variation is higher, then a higher budget is required to attain the same target availability level.

4. Three operational investment strategies

In this section, we consider three operational investment strategies for the resupply problem, one of which reflects the way in which the annual budget is currently spent at the RNN. These strategies are described in the Sections 4.1–4.3. In the last section, these strategies are compared for Example 1 on the basis of simulation results.

4.1. *Strategy 1: the currently used strategy*

Although currently, the RNN's inventory policy is focussed on meeting given service levels for individual products instead of on maximizing the availability of systems, we can still imitate the straightforward way in which the naval organization deals with the limited budgets in the exploitation period. The RNN divides the annual budget B into 12 equal parts, and at the beginning of each month one part is released. This leads to the following investment strategy, which we refer as the *Currently Used* (*CU*) *strategy*.

Each year y consists of 12 months of equal length. At the beginning of the *first* month, an amount of money equal to $B/12$ becomes available. Then one of the following three situations occurs:

(i) The inventory positions of all products $i = 1, \ldots, I$ are already equal to their maximum levels S_i^P.

(ii) The inventory positions are not equal to the maximum levels S_i^P for all products $i = 1, \ldots, I$, but the available money is sufficient to increase all inventory positions to the maximum levels S_i^P.

(iii) The inventory positions are not equal to the maximum levels S_i^P for all products $i = 1, \ldots, I$, and the available money is not sufficient to increase all inventory positions to the maximum levels S_i^P.

In situation (i), no immediate investments are made at the beginning of the month, and during the month all available money can be used to keep the inventory positions at their maximum levels as long as possible. The latter means that during the month, each item used from stock to replace a failed one, induces a purchasing order to replenish stock as long as this is allowed by the monthly budget. In the best case, there is enough money that the inventory positions remain at their maximum levels during the whole month. In situation (ii), at the beginning of the month one immediately orders new items to increase the inventory positions for all products to their maximum levels. Next, during the month the remaining money is used to keep the inventory positions at their maximum levels as long as possible. In situation (iii), all available money is spent at the beginning of the month (apart from a negligibly small amount of money that may remain because one can only procure whole items). Here, the first action is that for each of the products i with a negative inventory position the inventory position is increased to zero by ordering new items. Next, for $p = 1, 2, \ldots$, and for all products i for which the inventory position is smaller than S_i^p, the inventory positions are increased to S_i^p. This is continued until the whole budget has been spent. So, at this point, the solutions $\mathbf{S}^p$ of the initial supply problem are used. The motivation behind this rule is that it will lead to a close-to-maximal value for the current IP-based availability under the given circumstances.

At the beginning of the k-th month, $k = 2, \ldots, 12$, an amount of money equal to $B/12$ also becomes available. To this amount of money, one may add any money that possibly remains from the previous month. This amount is spent according to the same rules as described above for the first month. Notice, that the only difference between the first month and each of the other months lies in the fact that in the first month one can not use the money that possibly has been left from the previous month, since that money belongs to the previous years budget and hence it is lost. So, the available money at the beginning of each first month of a year is equal to $B/12$, while in the other months it may be more.

The CU strategy is such that at the beginning of each month the inventory positions are as far as possible returned to the balance point. Hence, the CU strategy is expected to lead to high end-of-year availabilities $A_{\text{eoy}}^{\text{IP}}(y)$.

In Fig. 4, we have depicted the behavior of the (conditioned) IP-based availability under the CU strategy for

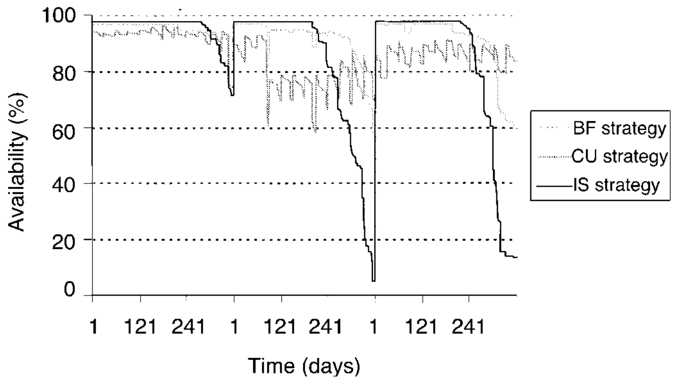

Fig. 4. Behavior of the IP-based availability over three consecutive years for Example 1 with given sample paths for the failures of items of the products $i = 1, \dots, I$ and under three different investment strategies.

Example 1 (with $B = 1.05 \times E\{\underline{D}\}$ and) with *given sample paths for the failures of items of the products* $i = 1, \dots, I$. The latter implies that the depicted IP-based availability is not equal to $A^{\mathrm{IP}}(y,t) = E\{A(\underline{\mathbf{X}}(y,t))\}$, but equal to $A(\mathbf{x}(y,t))$, where $\mathbf{x}(y,t)$ denotes the behavior that is obtained for the inventory positions of the products $i = 1, \dots, I$ as a result of the given sample paths for the failures and the CU strategy. The figure gives the behavior over three consecutive years of the exploitation period; namely years 15, 16 and 17, and in addition to the typical behavior that is obtained for the CU strategy, it also gives the typical behavior of the two strategies that are described in the next two sections.

4.2. *Strategy 2: the immediately spending strategy*

Under the CU strategy, the budget B is spent approximately equally over the year, however it may be better to procure, for as long as possible, each item immediately when needed to replenish stock. The latter may lead to higher, or at least equally high, inventory positions for (almost) all products during the whole year and thus to a better average IP-based availability. This idea is exploited in the second strategy, which is called the *Immediately Spending* (IS) *strategy*.

Under the IS strategy, at the beginning of each year y, the whole budget B becomes available, and then the situations mentioned in Section 4.1 as the situations (i)–(iii) may occur. Next, the same actions are taken as described for the first month of the CU strategy (see the third paragraph of Section 4.1), except that one now has to deal with a whole year instead of only the first month. In fact, the CU strategy can be described as a strategy under which the year is divided into $K = 12$ periods with equal lengths, and where an amount of money equal to B/K is released at the beginning of each period. Then, the IS strategy is identical to the CU strategy but with $K = 1$ instead of $K = 12$.

Under the IS strategy, the IP-based availability will usually reach its maximum level during the first part of the year. This is confirmed by Fig. 4. However, a main disadvantage of the IS strategy is that the end-of-year availability may become dramatically low (see Fig. 4). This happens when at the whole budget is spent at a relatively early point in the year. Then there is a high risk of seriously unbalanced inventory positions at the end of the year.

4.3. *Strategy 3: The balance focussed strategy*

The third strategy that we consider may be seen as a combination of the first two strategies. As under the CU strategy, each year y is divided into $K = 12$ periods with equal lengths. These periods are numbered by $k = 1, \dots, 12$. As under the IS strategy, the whole budget B is released at the beginning of the year. However, control parameters M_i^k are introduced to arrive at (almost) balanced inventory positions at the end of the year. Because of the latter implicit objective, the third strategy is called the *Balance Focussed* (BF) *strategy*.

Under the BF strategy, the following considerations are made in each period $k = 1, \dots, 12$ of a given year y of the exploitation period. The parameter $t_k = (k-1)/12$ is used to denote the beginning of period k, the actual inventory positions at this time instant are denoted by the parameters $x_i(y, t_k)$, $i = 1, \dots, I$, and the amount of money that is still available at this time instant is denoted by $B(y, t_k)$; so, for $k = 1$, $t_k = 0$ and $B(y, t_k) = B$. In principle, just as under the CU strategy, at the beginning of each period all inventory positions are increased to their maximum levels S_i^P (c.f., the situations (i) and (ii) as described in Section 4.1 for the CU strategy) or as high as possible if these levels can not be reached for all i (c.f., situation (iii)). Next, if it has been possible to increase all inventory positions to their maximum levels S_i^P, then these levels are maintained as long as possible. However, the following additional rule is applied. The number of new items of product i ordered at the beginning of, and during period k, may not exceed the control parameters M_i^k, $i = 1, \dots, I$. We now discuss how to set these control parameters M_i^k, at the beginning of each period k.

Whether the control parameters M_i^k are really limiting the procurement of new items, depends on the extent of the available budget $B(y, t_k)$ for the rest of the year in comparison to the amount of money that (in expectation) is needed to have the inventory positions at their maximum levels S_i^P during the rest of the year. For the latter, it is required that $S_i^P - x_i(y, t_k)$ new items of product i are procured at time instant (y, t_k) and in expectation $m_i(1 - t_k)$ new items of product i have to be procured during the rest of the year. Thus for all products together the following amount of money is expected to be needed:

$$\beta(y, t_k) = \sum_{i=1}^{I} c_i\left[(S_i^P - x_i(y, t_k)) + m_i(1 - t_k)\right].$$

If $B(y,t_k) \geq \beta(y,t_k)$, then there is no reason to limit the procurement of new items in period k. So, in that case the M_i^k are set equal to $M_i^k = \infty$ for all $i = 1, \ldots, I$.

Now suppose that $B(y,t_k) < \beta(y,t_k)$. Then the expected shortfall in money at the end of the year equals $\beta(y,t_k) - B(y,t_k)$. As a consequence, we foresee that at the end of the year the total investment in spare parts will be equal to (about) $C_P - (\beta(y,t_k) - B(y,t_k))$ instead of C_P. Thus, the inventory positions that we may have at best at the end of the year are given by the vector $\mathbf{S}^p$ where p is the largest index for which

$$C_p \leq C_P - (\beta(y,t_k) - B(y,t_k)), \tag{14}$$

(and $p = 1$ if this inequality is not satisfied for any index p). This inequality can also be written as

$$\begin{aligned} B(y,t_k) &\geq \beta(y,t_k) - (C_P - C_p), \\ &= \sum_{i=1}^{I} c_i\left[(S_i^P - x_i(y,t_k)) + m_i(1-t_k)\right] \\ &\quad - \sum_{i=1}^{I} c_i\left[S_i^P - S_i^p\right] \\ &= \sum_{i=1}^{I} c_i[S_i^p - x_i(y,t_k) + m_i(1-t_k)]. \end{aligned}$$

In the latter sum, the term between the squared brackets represents the number of new items of product i that should be procured during the rest of the year. For one or more products i, the current inventory position $x_i(y,t_k)$ might be sufficiently high that the i-th term of the latter sum is negative. In that case, we foresee that we end at higher levels than the desired levels S_i^p for the inventory position of these products i at the end of the year. This would imply that for the other products j the desired levels S_j^p can not be reached because of a lack of money. Hence, in that case, the target levels S_i^p must be slightly corrected by choosing p as the largest index p for which the following inequality is satisfied (instead of (14)):

$$B(y,t_k) \geq \sum_{i=1}^{I} c_i(S_i^p - x_i(y,t_k) + m_i(1-t_k))^+,$$

(and again $p = 1$ if this inequality is not satisfied for any index p), where $x^+ = \max\{0, x\}$ for all $x \in \mathbb{R}$. This leads to the following definition of the control parameters M_i^k:

$$M_i^k = \text{Round}\left((S_i^p - x_i(y,t_k) + m_i(1-t_k))^+\right), \quad i = 1, \ldots, I,$$

where Round(x) denotes the standard rounded value for each $x \in \mathbb{R}$.

This almost completes the description of the BF strategy. One last remark concerns the fact that under the BF strategy a positive amount of money may be left at the end of the year while not all inventory positions are equal to their maximum levels. This may be due to the role of the M_i^k and is likely to occur if in the last period less failures have occurred than expected. If such a positive amount of money is left, then this money is spent according to the same rules as described for situation (iii) in Section 4.1.

Notice that the control parameters M_i^k denote limits for the number of new items that may be procured for the products i during the rest of the year, i.e., during the periods $k, \ldots, 12$, while these parameters are only operational during period k. In general, in the first periods of each year, the parameters M_i^k barely limit the procurement of new items, even when they do not have finite values (i.e., when $B(y,t_k) < \beta(y,t_k)$). However as the end of the year approaches, they start to play their role. Hence, in comparison to the IS strategy, under the BF strategy the IP-based availability will have its maximum level somewhat earlier in the year, but a better behavior may be expected later on in the year. This is clearly demonstrated in Fig. 4.

4.4. *Numerical results*

In this section, the three strategies are compared on the basis of numerical results. Instead of presenting results for a whole series of examples, we restrict ourselves to the results for Example 1, for various reasons. First of all, it allows us to treat Example 1 extensively. Further, this example seems typical for systems occurring in practice (as it originates from the RNN), and from results for this example we may learn more than from other examples with artificial data. Third, our main goal is to assess the differences between the three strategies, rather than to determine for all kinds of cases which strategy is the best.

Example 1: A fire extinguishing system (continued)

For Example 1 with a budget $B = 1.05 \times E\{\underline{D}\}$, we have determined by simulation the average availabilities $A_{\text{avg}}^{\text{IP}}(y)$ and the end-of-year availabilities $A_{\text{eoy}}^{\text{IP}}(y)$ over the first 30 years of the exploitation period for all three strategies; see the results in Figs. 5 and 6, respectively.

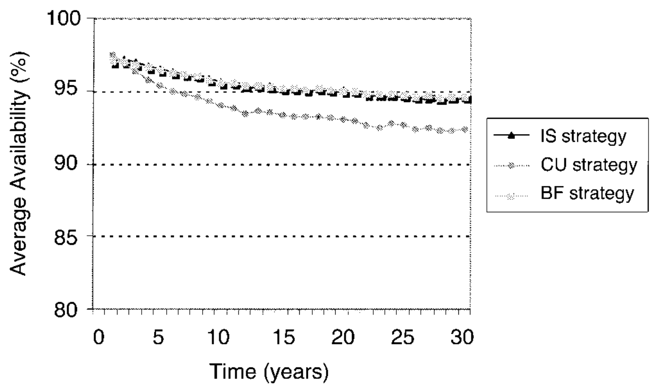

Fig. 5. The IP-based average availabilities $A_{\text{avg}}^{\text{IP}}(y)$, $y = 1, \ldots, 30$, for Example 1 with $B = 1.05 \times E\{\underline{D}\}$ and under three different investment strategies.

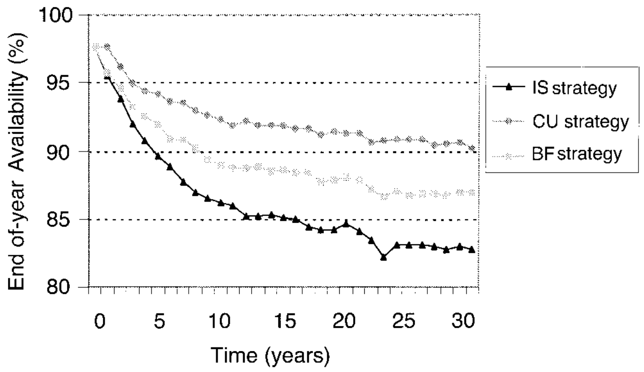

Fig. 6. The IP-based end-of-year availabilities $A_{\text{eoy}}^{\text{IP}}(y)$, $y = 0, 1, \ldots, 30$, for Example 1 with $B = 1.05 \times E\{\underline{D}\}$ and under three different investment strategies.

Table 2. The limit values for the IP-based average availabilities $A_{\text{avg}}^{\text{IP}}(y)$ and the IP-based end-of-year availabilities $A_{\text{eoy}}^{\text{IP}}(y)$ for Example 1 under three different investment strategies and with budget $B = aE\{\underline{D}\}$ and varying values for a

a	*CU strategy*		*IS strategy*		*BF strategy*	
	$A_{\text{avg}}^{\text{IP}}(\infty)$	$A_{\text{eoy}}^{\text{IP}}(\infty)$	$A_{\text{avg}}^{\text{IP}}(\infty)$	$A_{\text{eoy}}^{\text{IP}}(\infty)$	$A_{\text{avg}}^{\text{IP}}(\infty)$	$A_{\text{eoy}}^{\text{IP}}(\infty)$
1.01	47.0	44.2	52.9	28.7	54.1	35.6
1.02	70.4	67.1	78.9	52.7	79.9	61.0
1.03	83.5	80.8	87.7	66.4	88.9	74.1
1.04	89.4	87.5	92.3	77.1	92.2	81.5
1.05	91.1	89.5	93.5	81.1	94.0	84.9
1.07	94.2	92.8	95.6	88.1	95.8	91.0
1.10	95.8	94.6	96.7	92.9	96.7	93.8
1.15	97.0	96.3	97.5	95.7	97.4	96.1

For each strategy, the simulation was continued until for both $A_{\text{avg}}^{\text{IP}}(30)$ and $A_{\text{eoy}}^{\text{IP}}(30)$ the half-width of the corresponding 95%-confidence interval was equal to or less than 1.0% (in absolute value), and further for all three strategies the same sample paths for the failures were used. A computation time of about 20 minutes was needed (on a Pentium II, 266 MHz) for each strategy.

The results confirm our expectations. When comparing the strategies in terms of average IP-based availability, the current strategy CU is ruled out by both the IS and BF strategies. Next, comparing the two dominant strategies on the secondary criterion $A_{\text{eoy}}^{\text{IP}}(y)$, the BF strategy appears to be a clear winner. With respect to comparing the BF strategy to the CU strategy, it must be realized that under the BF strategy the behavior of the IP-based availability shows only one valley per year and lasts only a short amount of time (and its lowest point is given by $A_{\text{eoy}}^{\text{IP}}(y)$), as opposed to the CU strategy which may show twelve valleys per year (where $A_{\text{eoy}}^{\text{IP}}(y)$ denotes the lowest point of the last valley, not necessarily the lowest point of the curve over the year); see also Fig. 4. For that reason, the average availability of the CU strategy is significantly worse, while furthermore this observation indicates that comparing the CU and BF strategies in terms of end-of-year availability does not make any sense.

The curves for the end-of-year availabilities $A_{\text{eoy}}^{\text{IP}}(y)$ in Fig. 6 may be compared to the upper bounds given by the curve corresponding to "$B = 1.05 \times E\{\underline{D}\}$" in Fig. 3. This shows that the upper bounds are reasonable for the CU strategy, which is due to the small imbalance for the inventory positions that is obtained for this strategy at the end of each year. For the BF and IS strategies a greater imbalance is obtained at the end of each year and for them the difference between their end-of-year availabilities and the upper bounds is much larger.

The results in Figs. 5 and 6 were obtained for one particular choice of the budget B, namely $B = aE\{\underline{D}\}$ with $a = 1.05$. For other values of a (with $a > 1$) similar figures have been obtained. To show the impact of the budget B on the level of the availabilities, Table 2 displays the limit values $A_{\text{avg}}^{\text{IP}}(\infty)$ and $A_{\text{eoy}}^{\text{IP}}(\infty)$. Again, these results were obtained by simulation. For each combination of a value for a and a strategy, the simulation was continued until for both $A_{\text{avg}}^{\text{IP}}(\infty)$ and for $A_{\text{eoy}}^{\text{IP}}(\infty)$ until the half-width of the corresponding 95%-confidence interval was equal to less than 1.0% (in absolute value). The computation times varied from about 1 minute (for the cases with $a = 1.15$) to about 3 hours (for the cases $a = 1.01$).

From this table, the same conclusions can be drawn as from Figs. 5 and 6. In addition, observe that the differences increase when the budget becomes more tight (i.e., for decreasing values of a).

5. Conclusions and a preview of further research

In this paper, we have investigated how to optimize the operational availability of technical systems, consisting of a number of products, each of which are subject to failure. We have assumed that any failed item is condemned and should be replaced by a new one. Since procurement lead times may be significant, stocks of spare items have to be installed. The initial supply problem as studied by Sherbrooke (1992) assumes no condemnation and therefore concentrates on the determination of base stock levels, given an initial supply budget. Under the assumption of total condemnation, resupply budgets are needed to keep stocks and hence a system's availability at a desired target level. Based on an aggregate analysis, we have derived an upper bound for the end-of-year IP-based availability, which may be exploited to get a first idea on the annual budget needed. In addition, three operational investment strategies have been described and compared on the basis of numerical results for one example that originates from the RNN. The strategy that may be seen as a combination of the current way in which the people at the RNN deal with the limited annual budget and a straightforward approach, appears to be the best one.

The scope of the study presented in this paper is still limited. However, we would like to point out some important consequences. As mentioned already in the introduction, there are many reasons why either the system behavior is not stationary (as assumed in most stochastic models) or why external conditions may change. Due to technical progress, or insights directly obtained from earlier maintenance periods, the reliability of components may increase, or they may be replaced by more expensive, but also more reliable components. Because of this reason, but possibly also because of an attempt to decrease organisational costs, an organisation may be confronted with declining resupply budgets. In all cases it is extremely important to decide on how to spend a limited budget in order to maintain an optimal (within the new budget constraints) operational availability. Even more, models like the ones presented here may also be used as a tool to select a new system, taking into account not only initial purchasing costs but also operational costs, among which the costs needed for resupply of spare parts.

Future research will concentrate on the extension of the model and strategies presented here to multi-echelon, multi-indenture models and to the case with possibly different procurement lead times for the products (in which case the behavior of the actual availability and the behavior of the IP-based availability are not the same anymore). In addition, we will investigate situations where a failed item can either be repaired or be disposed of and replaced by a new one (basically, repair may be viewed as an alternative form of procurement, usually with lower lead times and costs than 'real' procurement; however, when modeling repair capacity explicitly, additional problems arise). In the latter models, both the costs of repair and of procurement will be taken into account, in order to obtain a clear view of the costs of a technical system during its operational life time. Such insights are believed to be extremely valuable in judging investments in alternative systems, by not only comparing initial investments but also operational life cycle costs, and therefore may contribute significantly to the development of life cycle cost models for complex technical systems.

References

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

Muckstadt, J. (1973) A model for a multi-item, multi-echelon, multi-indenture inventory system. *Management Science*, **20**, 472–481.

Rustenburg, W.D., Van Houtum, G.J. and Zijm, W.H.M. (1998) Spare parts management at the Royal Netherlands Navy: VARI-METRIC and beyond (revised version). BETA Working Paper WP-39, University of Twente.

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

Sherbrooke, C.C. (1986) VARI-METRIC: improved approximations for multi-indenture, multi-echelon availability models. *Operations Research*, **34**, 311–319.

Sherbrooke, C.C. (1992) *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques*, Wiley, New York.

Silver, E.A., Pyke, D.F. and Peterson, R. (1998) *Inventory Management and Production Planning and Scheduling*, Wiley, New York.

Slay, F.M. (1984) VARI-METRIC: an approach to modelling multi-echelon resupply when the demand process is Poisson with a gamma prior. Report AF301-3, Logistics Management Institute, Washington D.C.

Van Houtum, G.J. and Zijm, W.H.M. (1997) Incomplete convolutions in production and inventory models. *OR Spektrum*, **19**, 97–107.

Van Houtum, G.J. and Zijm, W.H.M. (1991) Computational procedures for stochastic multi-echelon production systems. *International Journal of Production Economics*, **23**, 223–237.

Biographies

W.D. Rustenburg (Jan Willem) is a Naval Officer in the Netherlands Navy. After the naval training-program, he served a couple of years as a deputy main engineering officer, and in the meantime he earned a M.Sc. degree at the Faculty of Mechanical Engineering of the University of Twente. Later on he was involved in a large re-engineering project at the Naval Establishment. His project activities were combined with Ph.D. research, in which he primarily focuses on spare parts management for complex technical systems. His project activities include the development and implementation of practical inventory control policies, redesign of distribution structures and development of purchasing strategies. He expects to complete his Ph.D. project and to successfully defend his Ph.D. thesis at the Eindhoven University of Technology in the Fall of 2000.

G.J. van Houtum (Geert-Jan) is an Assistant Professor of Operations Management at the Faculty of Technology Management of the Eindhoven University of Technology. He received his M.Sc. and Ph.D. degree of the Faculty of Mathematics and Computing Science of the Eindhoven University of Technology in 1990 and 1995, respectively. After his Ph.D project he worked for 4 years as an Assistant Professor at the Faculty of Mechanical Engineering of the University of Twente. In this period he was much attracted by and involved in production and inventory control projects at several companies in The Netherlands. At the beginning of 1999, he returned to Eindhoven, where his research is also inspired by practical problems. His current research interests include multi-echelon systems, spare parts management, performance analysis of manufacturing systems, design and analysis of materials handling systems, and E-commerce and its influence on supply chains.

W.H.M. Zijm (Henk) is a Full Professor in Production and Operations Management at the Faculty of Technology Management of the Eindhoven University of Technology. Previously, he has served among other things as a Full Professor at Twente University and as a senior consultant in Production and Logistics at Philips Electronics. He holds a M.Sc. degree from the University of Amsterdam and a Ph.D. degree from the Eindhoven University of Technology. His current research interests include manufacturing systems design, capacity planning and lead time management, shop floor control and process planning, design and control of warehousing and materials handling systems, project management and maintenance management. He has conducted numerous research projects with many industrial organizations and has published more than 80 papers in international scientific journals.