Integrating reliability centered maintenance and spare parts stock control

Willem van Jaarsveld* and Rommert Dekker

Econometric institute, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract

In the classical approach to determine how many spare parts to stock, the spare parts shortage costs or the minimum fill rate are a key factor. A difficulty with this approach lies in the estimation of these shortage costs or the determination of appropriate minimum fill rates. In an attempt to overcome this problem, we propose to use the data gathered in reliability centered maintenance studies to determine shortage costs. We discuss benefits of this approach. At the same time, the approach gives rise to complications, as the RCM study determines downtime costs of the underlying equipment, which have a complex relation with the shortage cost for spare parts in case multiple pieces of equipment have different downtime costs. A further complication is redundancy in the equipment. We develop a framework that enables the modelling of these more complicated systems. We propose an approximative, analytic method based on the model that can be used to determine minimum stock quantities in case of redundancy and multiple systems. In a quantitative study we show that the method performs well. Moreover, we show that including redundancy information in the stocking decision gives significant cost benefits.

1   Introduction

Availability of spare parts is important for companies, because spares are needed for efficient operation of capital goods. When equipment breaks down, the downtime can be significantly reduced if all spares needed for the repair are immediately available. If on the other hand spares are not immediately available, the waiting time for the spares can cause costly production losses. Because the costs of keeping spare parts on stock can be high, it is not obvious whether we should keep stock - either how many - to avoid downtime, or whether we should refrain from keeping stock to avoid holding costs. It is apparent from overviews of spare parts inventory control [1, 2] that most models aiming to support inventory decisions assume that certain pieces of information regarding the spare parts are available. Such pieces of information include the price and leadtime of the spare part, the usage of the part, and the shortage costs that are incurred during the waiting time for the part. The price and the leadtime of the part are in general available to the inventory

controller. The shortage costs and, in cases without demand history, the usage, are hard to estimate in practice. A method to circumvent the former problem is the setting of so-called service level targets, but finding appropriate values for these targets may prove difficult as well.

The research we report on was performed at a large petrochemical company. When determining stock quantities, obtaining reasonable estimates for the shortage costs was troublesome because of lacking data.

The company carries out reliability centered maintenance studies in order to improve maintenance practice at their plants. Reliability centered maintenance is a structured approach to ensure that all available data and knowledge is used to arrive at an optimal maintenance regime. As part of the particular type of RCM study carried out by the company the production loss incurred during equipment downtime and the estimated frequency of occurrence of different failure modes are quantitatively determined. The RCM study thus enables the use of downtime costs that are based on a thorough analysis by people that are actually working with the equipment.

We however find that while in inventory models often shortage costs consisting of a single number are assumed, in practice all equipment in which the spare part is used is a potential source of downtime costs, and the downtime costs of different pieces of equipment need not be equal. Another complication that came forward is redundancy. When there are two pieces of equipment, of which only one is needed to keep the plant running, a breakdown of one does not necessarily have severe economic consequences. In summary, it is not clear in what manner the downtime costs translate to shortage costs for the spare parts.

We contribute by proposing a new, versatile inventory model that can be used to tackle the above-mentioned complications resulting from the use of RCM in inventory control. While using RCM data for spare part inventory control has to our best knowledge not been described in literature before, there are some contributions on spare parts inventory control for redundant systems. In De Smidt-Destombes et al. [3] the tradeoff between repair capacity and spare part inventory control is investigated for a single k out of N system under condition based maintenance; i.e. when the number of defect pieces of equipment exceeds some previously defined limit, maintenance is initiated. They propose exact and approximate methods to analyze the system availability. In a later paper [4] the possibility that pieces of equipment degrade before failing is included, which complicates the analysis significantly and allows for more refined policies.

De Smidt-Destombes et al. [5] consider M identical k out of N systems under block replacement. For each system all defect pieces of equipment are replaced every fixed time interval. Two methods are proposed to analyze the system availability as a function of the number of spare parts stocked and the block replacement interval. Their most recent contribution [6] on the subject considers the optimization of the control parameters in the models presented earlier [3, 4, 5] to reach the target availability at minimal cost.

Chakravarthy and Gómez-Corral [7] also consider a single k out of N system, spare pieces of equipment, and a single repair man. A matrix analytic approach is used to evaluate the performance of these systems. Their model distinguishes itself from the models considered by De Smidt-Destombes et al. because they assume that spare parts are not requested when a piece of equipment fails with some positive, exogenous, state-dependent probability. Otherwise, their assumptions regarding the repair process are more restrictive than the assumptions investigated by De Smidt-Destombes et al.

Our model differs significantly from the models mentioned above. Neither the model we introduce, nor the existing models mentioned above are more general. The differences

between the models result from a difference in application. In the application examples given for the studies by De Smidt-Destombes et al., initiating maintenance involves a major setup cost and a significant setup time, elements that are both incorporated in their model. In discussion with the company at which this research was performed, we concluded that neither a setup cost nor a setup time play a significant role for our application. Both were not included in our model. Conversely, while other contributions only consider a single system [3, 4, 7] or multiple identical systems [5, 6], our model is very flexible in the sense that it allows an arbitrary combination of redundant systems, between which both the failure rate and the amount of redundancy may vary. The flexibility is needed to make the model applicable because many practical cases involve combinations of redundant systems with different redundancy levels and failure rates. Finally, our model is specifically designed to work with a detailed cost structure. It is therefore possible to model a system in which the throughput depends on the number of defect pieces of equipment in a gradual manner, another feature that is needed to make the model applicable for use with data coming from an RCM study.

Another related work is the paper by Dekker and Plasmeijer [8]. They advocate setting quantitative estimates for unit downtime costs in complex systems in order to facilitate decision making both on maintenance and on spare parts inventory levels. They provide methods to estimate these downtime costs. We take a different perspective. We will not estimate the downtime costs of individual pieces of equipment but instead directly estimate the shortage costs of spares in the combined system.

Besides proposing an inventory model specifically aimed towards working with RCM data, we contribute the following. In order to avoid the drawbacks of Monte Carlo simulation, such as long computation times and lack of insight, we propose two analytic approximations of the downtime costs. As a result of the difference in models, the approximations we propose are totally different from the approximations developed by De Smidt-Destombes et al. We develop an algorithm that can be used to determine the optimal reorder points based on the approximative methods. In a numerical experiment, we show that the cost increase as a result of using one of the approximations is small, while the other method has some performance issues, but is more intuitive to grasp and easier to implement. Furthermore, we give a number of qualitative arguments for the use of RCM data in inventory control. Finally, we present quantitative evidence that the value of using the detailed RCM data is significant. In particular, we compare the costs of the proposed methods with the costs of more traditional methods, and find significant cost benefits of the former over the latter.

This work expands on an earlier contribution (Van Jaarsveld and Dekker [9]). In the present work, we propose a new approximation with a significantly improved performance. We also add an extensive numerical experiment, from which many insights are derived.

The remainder of this paper is organized as follows. In Section 2 we give a problem setting including a motivation for the use of RCM data in inventory control. We also discuss the requirements of the model in terms of functionality and applicability. In Section 3 we give a formal description of the model itself. We discuss the practical issues that were taken into consideration when designing the model. In Section 4 a method is proposed to approximate the downtime costs using the model. In Section 5 we describe the setup of a numerical study. In Section 6 we give the results of this study, including investigations of the quality of the approximation and the benefits of using redundancy information over more traditional approaches. In the last section, we formulate conclusions.

2 Problem setting

2.1 Motivation for using RCM data for inventory control

Spare parts inventory control and reliability centered maintenance are related, in that the ultimate aim of both is to enable the efficient operation of equipment. Even so, the perspective taken differs between RCM and spare parts inventory control. Below we will describe RCM and inventory control while highlighting this difference in perspective, and we motivate the use of RCM data in inventory control.

From the perspective of spare parts inventory control, maintenance is primarily the source of the spare part demand. The specific maintenance requirements are kept more or less abstract, depending on the precise assumptions in each particular inventory model. Some contributions to inventory control assume the requirements of maintenance are translated to fill rate requirements on the overall or individual availability of parts. Others take the perspective that shortage costs are known. In the spare parts literature, little research is devoted to the actual determination of appropriate service targets and/or shortage costs. Notwithstanding, we run the risk of managing the inventory haphazardly in spite of the use of sophisticated inventory models if a well-defined procedure to determine the shortage costs is lacking, because without such a procedure we rely on parameters that cannot be managed nor controlled.

The influence of the availability of spare parts on the ability to perform efficient operation may however be quite complex. In certain environments, detailed knowledge of the equipment being maintained is needed to define appropriate availability targets for the spare parts. At the company at which this research was performed, a single spare part may be used in different pieces of equipment, each having different redundancy and different criticality. In that case, knowledge of redundancy and criticality is crucial to arrive at appropriate stock quantities.

From the perspective of RCM, maintenance is used to ensure that equipment continues to fulfill its intended function (see e.g. Moubray [10]). In order to successfully perform a reliability centered maintenance study, a thorough knowledge of the equipment to be maintained is indispensable. The team performing a RCM study contains people that are actually working with the equipment, with a lot of affinity with the equipment and a good knowledge of the consequences of downtime. Because of this knowledge, the team performing the RCM study is able to provide data of good quality on the downtime costs of the equipment, as well as data on the failure rates and redundancy structure of the equipment. This data is valuable for determining stock quantities. Furthermore, this data is delivered in well-structured reports, with standard data fields that allow the use of (semi-)automatic inventory models for a lot of different parts. And even though different RCM studies may differ, the exact procedure of each study is often well-defined. These properties make that data coming from an RCM study is particularly valuable when managing spare parts inventory, and they make RCM a good basis for a well-defined procedure to determine spare parts shortage costs.

In summary, since RCM data is well-structured and delivered by a team with a lot of knowledge of the equipment, and since it contains information directly related to the costs of waiting for spares, integrating RCM data in inventory control gives opportunities for cost effectively improving the inventory decisions. Moreover, integrating RCM in inventory control ensures that the inventory decisions are based on the best information and knowledge available to the company.

2.2 The company

This research is a result of a cooperation with a company that is active in the petrochemical industry. The company at which this case study was performed is stakeholder in large refineries and oil production platforms. RCM studies are performed on critical equipment for both greenfield and existing projects. Equipment assessed in the study includes feeding pumps, power generators, valves and other components that perform important functions. Equipment failure may lead to production loss, which can be translated to downtime costs. As discussed before, the target of the study is to incorporate the RCM data in inventory control to improve inventory decisions and to ensure that all information available to the company is used in the best way to ensure optimal inventory control.

The RCM study carried out by the company at which the research was performed is quantitatively oriented, quantitative estimates of the effect of equipment downtime and failure rates are estimated by the RCM team. This makes this type of RCM study particularly valuable for inventory control. Another aspect that comes forward from the study is redundancy. In particular, the study identifies redundancy and estimates the downtime costs as a function of the number of pieces of equipment functioning within the group of equipment for which the redundancy is involved. This information is crucial when determining spare parts shortage costs. At the same time, to be able to use this information, an inventory model is needed capable of using it.

Even though the research was performed in cooperation with a company in this particular industry, the developed model is quite general and can be applied across a wide range of industries, as long as reliable estimates of the required data are available.

2.3 Model requirements

The model we develop should enable the use of data coming from the RCM study for the purpose of inventory control. We should be able to apply the model cost-effectively and with good results for the majority of the equipment for which the RCM study is carried out. Not all data coming from the RCM study needs to be used. Aspects for which the added value of including them does not outweigh the decreased usability of the model because of the increase in complexity should be excluded. The primary reason to refrain from using complex models is the increased burden they put on the data collection. It must not be forgotten that RCM studies are primarily carried out in order to optimize maintenance, and that information needed for that purpose is not necessarily essential for stock control.

The balance between realism and applicability evidently depends on the specific application. For specific, very costly equipment with huge downtime costs ($\geq \$10^6$/day), a detailed simulation may be a cost-effective approach to determine appropriate stock quantities for very expensive spares ($\geq \$10^5$). We aim at spare parts with somewhat lower, but still considerable, costs, and high downtime costs. The cases considered in Section 5.2 give a good picture of the type of the applications for which the model was developed. In a single RCM study, hundreds to thousands of pieces of equipment may be considered. Because the time spent by a stock analyst is expensive, we should limit the effort required to find an appropriate stock quantity for each part. Therefore, the decision making process for these parts should be (semi-) automatic. When developing the model we must therefore limit the number of details included, because each included detail requires data to be collected. At the same time, we must make sure that the most important characteristics are included to ensure that the quality of the decision is sufficiently high.

3 The model

The model that resulted from the development is described in this section. The model assumes that failures can occur in multiple pieces of equipment in the plant, which need not have the same downtime costs. The model can also cope with redundancy. This structure enables the use of the model in combination with the data coming from the RCM study. In addition to giving a formal description of the model, we discuss the practical considerations that were taken into account when designing the model. In particular, we mention some aspects that were omitted from the model with the purpose of simplifying the data collection. The discussion of these aspects will be based upon the application for which this model was developed. It can be helpful when determining whether the model is suitable for use in another application.

3.1 The spare part

A spare part with deterministic leadtime L is given. The spare part is stocked in a single location, and it is used solely in the repair of some pieces of equipment. We assume that the part is controlled in a continuous review policy. We concentrate on expensive spare parts; we assume a base stock policy with base stock level S is used for the part. We will refer to this base stock level as a reorder point in the rest of this paper. Under a continuous review base stock policy a part is ordered each time it is withdrawn from stock (see e.g. Axsäter [11, pp. 49-50]). The total number of parts on order plus the number of parts on stock is kept constant, and equal to S. We assume that the holding cost per unit of time for a base stock policy S are given by hS, where h is the holding cost per part per unit of time. Often a fixed percentage of the value of the part is used for this. Note that we assume that holding costs are paid for parts in replenishment as well.

We mention an important issue that needs to be resolved in order to apply the model. Data on leadtime and cost must be obtained from the supplier in a standard format. We will not go into details on this issue, but we mention that a well-defined procedure to obtain reliable data on the spare parts themselves is essential for successful inventory control. Furthermore, data must be obtained describing which parts are used in the repair of which equipment. Even though this may not be standardly included in an RCM study, this data can be obtained during the study. The data can also be obtained in a separate study.

3.2 Redundancy and downtime costs

We model the redundancy by assuming that the pieces of equipment are partitioned in N functional groups. Pieces of equipment in the same functional group perform the same function together involving redundancy. The number of equipment pieces in functional group $n \in N$ will be denoted by R_n. E.g. consider a case with 3 pumps. One pump performs a single function on its own, while two pumps perform another function together. Than, we have 2 functionality groups 1 and 2, with $R_1 = 1$ and $R_2 = 2$.

In each functional group, equipment may break down. We assume that as long as there are active pieces of equipment in functional group n, breakdowns occur in the functional group with rate λ_n. The probability of a failure in a functional group in some small time interval is thus assumed to be independent of the number of pieces of equipment that are down in the functional group (as long as at least one is still running), and independent of the time that the system has been running without failures.

A possible issue is that the failure rate in a functional group may in practice depend on the number of pieces of equipments that are running. E.g. for multiple pumps running in parallel in a two out of two (denoted as 2oo2 from now on) configuration, the failure of one pump may increase the stress on the other pump, which may increase the failure rate of the functional group. On the other hand, when only one pump is running, only this pump can possibly fail, which may in turn reduce the failure rate of the functional group. As stated above, we ignore these subtleties, and assume a failure rate that does not depend on the number of active pieces of equipment in the functional group. The main motivation for this simplification is that the estimation of the dependence of the failure rate on the number of running pieces of equipment requires a lot of additional effort during the RCM study and when preparing the data for the model. The return for this effort in terms of improved results is probably limited.

Depending on the number of pieces of equipment defect in functional group n, we incur downtime costs. For $i \in \{1, \ldots, R_n\}$ pieces of defective equipment in group n we assume that downtime costs of $c_{ni} \geq 0$ per time unit are incurred. These downtime costs will be estimated in the RCM study. Note that we are interested in the marginal costs of having a unit of extra downtime, the fixed repair costs should be excluded from c_{ni} since they are not affected by the number of spares.

The assumption of marginal downtime costs linear in the downtime need not always hold. In some cases, the downtime cost may depend non-linearly on the downtime interval because of pipeline capacity. Again and for the same reasons, we ignore this subtlety.

3.3 Repair process

When equipment breaks down, we assume a single spare part is needed. In case the part is available from stock, it is immediately withdrawn. If stock is depleted, parts are assigned on a first come first serve basis to the repairs. As soon as a part is available for the repair, the repair of the part is commenced. The repair time from that moment on is given by $t_{\bigtriangledown}$.

We acknowledge that first come first serve need not be optimal, because the downtime costs for different pieces of equipment are not equal. It may even be suboptimal to always withdraw stock when it is available. In order to keep stock for future, more critical, demands, some stock may be rationed for use in more critical equipment (see Kleijn and Dekker [12] for an introduction to inventory rationing and an overview of relevant literature).

The possibility of spares failing before being used or upon demand is not taken into account. Most suppliers ar not willing to share information regarding failures upon demand because of commercial reasons, and this data is hard to collect in another manner. For parts in which failure upon demand plays an important role the method is not suitable.

Cannibalization of parts from other, less critical equipment is excluded. The expediting of orders for spare parts when the spare parts are really needed is also excluded. Including these would require an estimate of the cost- and time- parameters related to these special operations, which would decrease the usability of the model.

The assumption that obtaining the spare part and repairing the equipment commence immediately after the equipment breaks down does not hold in general. Systems in which repairs on redundant systems are deliberately postponed are considered by De Smidt-Destombes et al. [3, 5]. Postponement of the repairs are in these cases justified by fixed setup costs. We consider systems in which repairing multiple pieces of equipment within the same functional group simultaneously does not give an economic advantage. Repairs should thus in principle be immediately commenced once they are detected. In redundant systems

however, a failure need not be immediately detectable. For instance in a 1oo2 system, the piece of equipment not running may degrade and break down, which happens undetected until one attempts to activate the equipment. Undetected failures are a big threat that should be mitigated by frequent testing, for instance by switching of the running equipment. If testing is so frequent that failures are detected within a time span much smaller than the other time scales of the problem, the applicability of the model is not hurt by the undetected failures.

A further limitation of the assumption of immediate repairs and constant repair times is finite repair capacity. When a lot of parts fail in a short time interval, there may not be sufficient repairmen to finish all repairs in regular working hours. Note however that the repairmen perform the maintenance for all equipment on the site, while the spare parts are only used in a very limited number of equipment. The relative variation in the workload of the repairmen is thus considerably less than the relative variation in the usage of the spare parts, as a result the capacity problem related to repairmen is relatively less costly to mitigate. Furthermore, to finish at least the repair of the critical equipment considered in this paper overtime should be considered. Especially for redundant systems one may however be tempted to postpone a repair rather than using overtime because there are no immediate consequences of delaying the repair. One should however always keep in mind that the system was designed as a redundant system primarily because it has high criticality, and using the redundancy to provide repair flexibility reduces the mitigating effects of the redundancy. In summary, we argue that the effects of finite repair capacity should be limited if the system is properly controlled.

4 Approximate analysis

The purpose of the model presented in the previous section is the optimization of the reorder point. In order to optimize the reorder point, we need to determine the downtime costs for different values of the reorder point. The amount of downtime incurred in a certain time interval depends on coincidental circumstances. We will make decisions based on the average value over all possible outcomes. We are thus interested in the long term expected downtime costs.

One method to obtain downtime costs is Monte-Carlo simulation. However, the systems that we are considering are rare event systems, especially for high reorder points and redundancy. We will show in Section 6 that using Monte Carlo simulation for optimization requires very long CPU times, which may hamper applicability. Furthermore, simulation does not give much insight into the problem.

To overcome the problem of long simulation times and to gain insight in the model, in this section we will propose two approximative methods for determining long term expected downtime costs. Both methods are based upon decoupling the warehouse process and the repair process for the different functionality groups.

In Section 4.1 we show that under deterministic waiting times for spares, the expected amount of downtime and the downtime costs can be easily evaluated. In Section 4.2 we use this observation to propose two approximative methods for evaluating the downtime costs for a given reorder point. In Section 4.3 we show how to use the approximations to determine reorder points. In Section 4.4 we present two traditional methods that will be used to benchmark our approach, and to gain insight in the relative value of using the more sophisticated approaches described in Sections 4.1-4.3.

4.1 The repair process

The total repair time is composed of the waiting time for spares and the remaining repair time after spares are available ($t_\bigtriangledown$). The waiting time for individual repairs depends on the state of the warehouse at the moment of failure. This couples the different functional groups, and gives the systems its complexity.

In this section, we show how to calculate the long term expected downtime costs for a single functional group under the assumption that the waiting time for spares, and thus the total repair time, is fixed and deterministic. While this assumption is not satisfied in our model, the theory developed in this section will serve as a building block for the approximations for the downtime cost of the whole system that will be presented in the next section. We denote the deterministic waiting time for spares by $t_{w,d}$. The total repair time is then also deterministic, and has length $t_\bigtriangledown + t_{w,d}$, which will be denoted by t_d.

Note that under the assumption of deterministic waiting times, the functionality groups decouple. Consider functionality group $n \in N$ (using the notation introduced in Section 3.2). This functionality group can be represented by a closed queueing network with R_n customers and two stations: (i) an ample server with mean service time t_d, representing the repair process; (ii) an exponential server with mean service time $1/\lambda_n$, which represents the failure process. This network belongs to the class of so-called BCMP networks and thus has a product-form solution (see Baskett et al. [13]). The steady state probabilities of having $i \in \{0, \ldots, R_n\}$ defect pieces of equipment in this functional group is thus equal to

$$p_n(i, t_d) = \left(\frac{(\lambda_n t_d)^i}{i!} \right) \bigg/ \left(\sum_{j=0}^{R_n} \frac{(\lambda_n t_d)^j}{j!} \right). \tag{1}$$

The long term expected downtime costs C_n for functional group n can be calculated from the steady state probabilities given by Eq. (1) using the following relation

$$C_n(t_d) = \sum_{i=1}^{R_n} c_{ni} p_n(i, t_d). \tag{2}$$

The total downtime costs are the sum of the downtime costs for the individual functional groups,

$$C_{\text{total}}(t_d) = \sum_{n \in N} C_n(t_d). \tag{3}$$

4.2 Approximating the downtime costs

In the previous section we showed that under the assumption of deterministic waiting times for spares, the downtime costs can be computed efficiently. In this section we will present *two methods* to approximate the total downtime costs for the dynamic system.

For *both approximations*, we will approximate the total demand rate for spares by a Poisson demand stream. We have assumed that the rate at which failures occur is given by λ_n for every functional group n, as long as there is any equipment running in that functional group. Each time a failure occurs, a spare part is needed. In practice, downtime of equipment is short in comparison to the uptime of the equipment even in case no spare parts are stocked. Each functional group thus gives rise to a demand stream which can be

approximated by a Poisson process with rate λ_n. It is thus reasonable to approximate the total demand rate for spare parts as a Poisson process with rate

$$\lambda = \sum_{n=1}^{N} \lambda_n. \tag{4}$$

The *average waiting time approximation* will be based upon the average waiting time for spares, which will be denoted by $\bar{t}_w$. It depends on the total demand rate (approximated by λ), the leadtime L, and the number of spare parts kept on stock S. It will be calculated analytically based on Little's formula:

$$\bar{t}_w = L - \frac{S}{\lambda} + \frac{1}{\lambda} \sum_{i=0}^{S-1} \frac{(S-i)(\lambda L)^i}{i!} e^{-\lambda L}. \tag{5}$$

The total average expected repair time is now given by

$$\bar{t} = \bar{t}_w + t_{\triangledown}. \tag{6}$$

We will use this average in Eq. 3 to approximate the downtime costs:

$$C_{\text{average}} = C_{\text{total}}(\bar{t}) \tag{7}$$

The approximation is similar to the one used in the analysis of the METRIC model, proposed by Sherbrooke [14]. In Section 6 we will see that the approximation sometimes performs poorly. It has the advantage of being easy to grasp and easy to implement.

To improve the performance we propose a second approximation, the *dynamic-static waiting time approximation*. The repair resulting from an arbitrary breakdown incurs a stochastic delay due to the waiting time for spares, which can be zero or positive. Under the approximative assumption of a Poisson demand stream for spares, a simple expression can be derived for the distribution of this stochastic delay. We will first derive this expression. Then we will show how to use the expression to obtain an approximation for the downtime costs.

Consider an arbitrary breakdown, which we assume to occur at time t. Recall that we assume a base stock policy with reorder point S is used and that stock is allocated on a first come first served basis. Consequently, the part that was ordered when the Sth breakdown preceding the current breakdown occurred, will be used in the repair of the current breakdown. Say this earlier breakdown occurred at time $t - X$. Recall the approximative assumption of a Poisson repair stream. Then X is Erlang-k distributed, with $k = S$: it is the sum of S exponentially distributed variables with mean $1/\lambda$. When $S = 0$ then $X = 0$ with probability 1. The part arrives at time $t - X + L$. Therefore, at time t the remaining waiting time is $\max(0, L - X)$. This random variable will be denoted by Y.

The *dynamic static waiting time approximation* is based on the approximative assumption that we may take the expectation of the deterministic waiting time expression of the downtime over the steady state waiting time distribution:

$$C_{\text{dyn.-st.}} = E(C_{\text{total}}(Y + t_{\triangledown})) \tag{8}$$

where C_{total} is determined using Eq. (3). Note that the *average waiting time approximation* differs because

$$C_{\text{average}} = C_{\text{total}}(\bar{t}) = C_{\text{total}}(E(Y) + t_{\triangledown})$$

Compared with simulation, it is relatively easy to evaluate Eq. (8) numerically, a number of standard methods are available. For completeness, we describe how we evaluated the expression in our numerical experiments. We use that

$$
\begin{aligned}
C_{\text{dyn.-st.}} =& E(C_{\text{total}}((L-X)^+ + t_\triangledown)), \\
=& P(X > L)C_{\text{total}}(t_\triangledown) + P(X < L) \\
& \times E(C_{\text{total}}(L - X + t_\triangledown)|X < L).
\end{aligned}
$$

Since an analytic expression for the CDF of the Erlang distribution is known, the first term can be easily calculated. The PDF of the Erlang distribution is also known, let's denote it by $f(x)$. We have that

$$
\begin{aligned}
& P(X < L)E(C_{\text{total}}(L - X + t_\triangledown)|X < L) \\
& = \int_0^L f(x)C_{\text{total}}(L - x + t_\triangledown)dx, \\
& \approx \frac{1}{N}\sum_{i=0}^{N-1} f(i/N)C_{\text{total}}(L(1 - i/N) + t_\triangledown).
\end{aligned}
$$

In our numerical tests, we use $N = 10^5$.

4.3 Optimization

We now show how to find S that minimizes the average waiting time approximated costs, and S that minimizes the dynamic-static waiting time approximated costs. The overall costs are given by

$$
hS + C_{\text{average/dyn.-st.}}(S) \tag{9}
$$

where C_{average} and $C_{\text{dyn.-st.}}$ are given by Eqs. (7) and (8), respectively. These optimizations are optimizations over a single, discrete parameter, which we can easily perform if we can determine a suitable upper bound.

An upper bound can be obtained if we make the reasonable assumption that the downtime costs of a functional group are nondecreasing with an increasing number of defective equipment ($i > j \rightarrow c_{ki} > c_{kj}$). In that case, it is clear that

$$
C_{\text{total}}(t_\triangledown) \tag{10}
$$

is a lower bound on the downtime costs, both when calculated using C_{average} and when using C_{static}. For later reference, note that it is also a lower bound for the true downtime costs. To optimize S we start by setting $S = 0$, and we determine the costs. We then repeatedly increase S by 1, and keep track of the solution with minimal total costs. It is not hard to see that the search can be terminated once a value is for S is reached for which $hS + C_{\text{total}}(t_\triangledown)$ exceeds the cost of the best solution found so far.

4.4 Traditional methods

We will also examine two methods for solving the problem, that will be used as a benchmark in the numerical experiment. The traditional methods do not take into account detailed redundancy information, and they represent what companies might do if data regarding

11

redundancy is not available or if they lack the know-how or organizational structure needed to couple the redundancy information with spare parts stock control.

The first method assumes that an estimate is used for the total demand rate (equal to λ, as given by Eq. (4)) and the leadtime. The system is modelled as a base stock model in which demand is back ordered. Then, it is easy to evaluate the fraction of parts that are delivered from stock in the steady state of the system, a fraction often referred to as fillrate (see e.g. Axsäter [11, pp. 94-95]).

In the method, we then use a fill rate target to determine the reorder point. The lowest reorder point for which the target fillrate is reached is chosen. This method assumes that there is no additional information available regarding the downtime costs and the redundancy structure. Therefore, in practice there would be no information on which we could base the setting of different fill rate targets for different parts. We will therefore assume that a fixed target is used for all cases. Notwithstanding, we wish to gain some insights into the dependency of the results on the overall fill rate target. Therefore, the *overall* fill rate target is varied to obtain different types of the fill rate target traditional method. We will denote these different types by their target fill rate in quotation marks, so *"95%"* denotes the method that uses a fillrate target of 95% over all cases.

In the second traditional method, we assume that the highest possible downtime cost for all pieces of equipment in which the part is installed can be determined. This downtime costs is then used as the penalty cost per time period in the system, again modelling the system as a Poisson demand system in which demand is back-ordered. The overall costs (downtime + holding) are then minimized to obtain the optimal reorder point. This method is similar to the method currently used by the company as a recommendation to the stock analysts (Trimp et al. [15]). Note that while it has some awareness of the downtime costs, it does not use any redundancy information. We will refer to this second traditional method as *benchmark method* later on.

5 Experimental setup

5.1 Simulation

The analysis discussed in the previous section gives approximative estimates of the downtime costs, based on which approximately optimal reorder points can be determined. In order to test the quality of these approximations and the resulting recommendations, we describe in this section a simulation approach that enables us to find asymptotically exact estimates of the downtime costs. We also describe the approach that was used to find the optimal reorder point using simulation.

To avoid confusion, let us first stress that in the simulation we aim to find downtime costs for the model described in Section 3. All assumptions discussed in that section thus remain in place. Our aim is to estimate the quality of the approximations and the traditional methods presented in Sections 4.2 and 4.4, and the quality of the recommendations that result from using the approximations.

In order to assess the effect of these assumptions we need the true costs of the system. To this end, we simulate the system. We use event-driven simulation. Events correspond to breakdowns, arrivals of orders for spare parts, and repairs finishing. The system is simulated for a long period of time. This time period is divided in a number of different batches with a length of 1000 years. To make sure that the costs in the different batches can be safely assumed to be independent estimators of the true costs, in between the batches

we keep periods of 100 years that are not used for statistics. Autocorrelation tests indicate that 100 years is much longer than the autocorrelation of any of the systems considered. For each of the 1000 year periods, the costs can be obtained. This allows us to determine asymptotically correct estimates of the downtime costs and the associated variances for an arbitrary instance of the model and a given reorder point.

To find the optimal reorder point using simulation, we would like to find the associated cost for all base stock levels up to some upper bound with very high precision. However, since the models correspond to rare-event systems, this is impossible for some cases. In order to obtain consistent results, the following procedure has been developed.

We start by considering $S = 0$, and increment S each step. For each reorder point, we simulate the system until we have 1000 periods of 1000 years for which the total downtime costs are *positive*. Based on the simulation period obtained in this way, we obtain an initial estimator of the downtime costs and its associated standard deviation. While for most systems in each 1000 year period downtime costs are incurred, in some systems in only 1 out of more than 250 periods positive costs are found. The described method is used to ensure that it is reasonable to apply the central limit theorem for the estimator. We continue increasing S until *both* the following conditions are met:

- The current reorder point exceeds the "optimal" reorder points calculated using the different approximative methods described in Sections 4.2 and 4.4.

- The holding cost for the current solution plus the lower bound for the downtime cost (see Section 4.3) exceed the sum of the estimate for the total costs for the best reorder point found so far plus five times its standard deviation. This last measure is used to ensure against cutting off the optimization prematurely because of statistical variation.

Using the above method we obtain an estimator for the downtime costs of each of the solutions. Note that we have taken extensive measures to assure that the optimal reorder point is included.

The relative deviation of the estimators may however be quite high at this point. In order to decrease the variance of the estimators, we continue by simulating for each reorder point until the resulting estimator reaches a target relative deviation. We set the target precision at $2^{-10} \approx 10^{-3}$. Note that there is no target precision for which we can guarantee that the reorder point with minimal estimated cost is the reorder point corresponding to the minimum true costs. Based on the central limit theorem is however reasonable to assume that the costs of deviations from the true minimum will not be much larger than a few times the target precision.

For some systems, reaching this target costs too much computation time, which forces termination if we want results at all. Therefore, we choose to gradually increase the target for each considered case. We start with a target of $2^{-0} = 1$. We then increase the precision by iteratively halving the target precision. We stop this procedure when a target precision of $2^{-10} \approx 10^{-3}$ is reached. When the simulation time for a reorder point exceeds a predefined limit on the simulation time, we will terminate the simulation for the case considered without reaching the target precision. However, we will still have reached a relatively consistent precision over all reorder points for that case. We set the limit at a simulation time of 10^6 periods of 1000 years. When the search is terminated before attaining the required precision, we obtain an estimate of the required additional CPU time that would be needed to attain the target. To obtain this estimate we use that the standard deviation of the

estimator is $\sim 1/\sqrt{n}$, where n is the number of 1000 year periods.

installed base descr.	functional group (n)	c_{n1}	c_{n2}	c_{n3}	λ_n
		($\times 365k\$$/yr)			(/yr)
business	1	4	-	-	0.5
case	2	0	30	-	0.66
	3	0	20	100	1
$5 \times 1oo2$	1	0	30	-	0.5
	2	0	30	-	0.5
	3	0	30	-	0.5
	4	0	30	-	0.5
	5	0	30	-	0.5
$2 \times 1oo2$	1	0	100	-	0.5
	2	0	100	-	0.5
$1oo1$	1	10	-	-	0.5
$1oo2$	1	0	20	-	0.66
$1oo3$	1	0	0	100	1
$2oo3$	1	0	40	100	1

Table 1: The configuration of the installed bases used in the numerical study.

parameter	uom	values
leadtime	$\times$ weeks	1/7, 1, 4, 8, 22, 52
holding costs	$\times k\$$/yr	0.125, 0.625, 2.325, 6.25
repair time	$\times$ weeks	1/7, 1, 6

Table 2: The other parameters that were varied in the numerical study.

5.2 Cases

We now describe the cases that are considered in the numerical study. We consider all combinations of a set of leadtimes, a set of repair times, a set of part costs, and a set of installed bases (a number of functional groups in which the part is used). We use 7 different possible settings for the installed bases, the values are given in Table 1. The values for the other parameters that are considered are given in Table 2. In total, 504 cases are considered.

In the design of the cases, we have taken into account some empirical knowledge that will be summarized in the following. The leadtime may vary significantly over different parts, and also over different locations. In some cases, the parts are made to order. In case of complex equipment, this may induce leadtimes of a year or even more. Other factors that may cause significant leadtimes are customs delay, and the fact that some equipment can only be moved using special transport. On the other hand, some parts may be obtained from a central warehouse in less than a day.

Repair times may be anything from a few hours to a number of weeks. In our discussion with the company, it became clear that it would be unlikely that a repair would take more than a few weeks if all spare parts that are needed are available.

The holding cost is generally fixed at 25% of the value of the spare part annually. For the value of the spare part, anything between a few dollars and $k\$200$ seems reasonable. We

however did not include spare parts of over $k\$25$, since for them we expect a more thorough analysis including some factors that were omitted from the model to be cost effective. We also did not take into account parts with a value less than $k\$0.5$ because it seems unlikely that a detailed study of the redundancy will be cost effective for them. Moreover, such low-value parts will be ordered in batches to reduce marginal ordering costs, which puts them out of scope of this research since our analysis is based on the assumption of a base stock policy.

For the functionality groups, it is hard to come up with a range that includes all possibilities. We have included some basic systems: 1oo1, 1oo2, 1oo3, 2oo3. We have also included three combinations of a number of these systems. In general, we let more redundant systems have higher associated downtime costs, because systems are often made redundant because they perform an important function. For the downtime costs, anything between a few $k\$$ per day and $1000k\$$ per day seems reasonable. However, we again do not include cases with very high downtime costs, because for them we expect a more detailed analysis.

6 Results & discussion

The results of the numerical experiments are presented and discussed in this section.

6.1 Computation times

As stated before, the main reason for using approximative methods to optimize the reorder point instead of simulation are the significant CPU times required for simulation optimization. The required processor times for the simulation optimization of the considered cases are shown in Table 3. The simulations were performed on a 2.33 GHz CPU dual core with 3.23 GB of RAM, on two separate threads running in parallel.

attained precision	# of cases	proc. time (avg. per case) performed	for target
2^{-10}	368	9 min	-
2^{-9}	44	34 min	2 hrs
2^{-8}	40	51 min	13 hrs
2^{-7}	28	30 min	32 hrs
2^{-6}	12	28 min	482 hrs
2^{-5}	7	22 min	1478 hrs
2^{-4}	5	22 min	5882 hrs

Table 3: The statistics regarding the attained precision and the CPU times in the simulation. We also tabulate an estimate of additional CPU time that would be needed to attain the precision target.

The target precision was set at a normalized standard deviation smaller than $2^{-10} \approx 10^{-3}$. For the cases for which the target was attained, the average simulation time was 9 minutes. Note that such a computation time may already severely hamper the applicability, since it will discourage practitioners to use the system as a decision support tool because they might not be willing to wait.

Some simulations were terminated before the target relative deviation was attained. Statistics regarding this point are also tabulated. An estimate of the additional simulation time that would be required to attain the target is tabulated as well. The table shows

that while the target precision is relatively modest, attaining it would require a prohibitive amount of effort for some cases. This limits the applicability of using simulation for optimization: if a reasonable precision cannot be attained for the different estimators of the downtime, the resulting reorder point could be far from the optimum.

The computation time of the other, approximative methods to determine the optimal reorder point was such that it will not be a practical issue. Of the approximative methods, the dynamic static waiting time approximation consumed the most CPU time, but it was still very fast. The average time it took to perform one optimization using this method was 70 ms, the maximum optimization time was 400 ms. Users will experience this as practically instantaneous, which is a major advantage of the approximate methods over using simulation to optimize the reorder points. Another advantage of the approximative methods is that it requires much less effort to implement them then to implement a simulation approximation approach.

6.2 Precision of downtime cost approximations

Another subject of interest is the performance of the different approximate models. We distinguish two types of performance: the precision of the estimates of the downtime costs, and the quality of the recommended reorder points.

The results regarding the precision of the approximations are shown in Table 4. The statistics are based on all 2443 considered case - reorder point combinations for which the target precision was reached. The 746 combinations for which the target precision was not reached were omitted from the statistics in the Table, since it is impossible to gauge the performance of the statistics when we cannot rely on the downtime estimate obtained using simulation. While some cases were not taken into consideration, the cases that were considered still represent a very broad range of systems. Therefore, the conclusions that we draw regarding the performance of the approximations have a strong predictive power for other instances of the model.

We distinguish three approximations: the two approximations developed in Section 4.2 ($C_{\text{dyn.-st.}}$ and C_{average}), and the second traditional method developed in Section 4.4 (which is denoted by C_{bmrk}). We cannot use the method based on the fill rate target at this point, since that method does not estimate downtime costs.

statistic	$C_{\text{dyn.-st.}}$	C_{average}	C_{bmrk}
$|(C_e - \mu)/\mu| < 1\%$	83.1%	73.4%	3.6%
$|(C_e - \mu)/\mu| < 5\%$	97.5%	78.5%	6.2%
$|(C_e - \mu)/\mu| < 10\%$	99.5%	82.8%	13.2%
$|(C_e - \mu)/\mu| < 50\%$	100%	95.0%	15.7%
$\max C_e/\mu$	1.14	1.27	1823
$\max \mu/C_e$	1.002	5.4	0.999
$\text{avg}|(C_e - \mu)/\mu|$	0.007	0.07	96

Table 4: Different statistics regarding the precision of the proposed methods. μ denotes the estimate of the downtime costs based on simulation, C_e denotes the approximated costs.

To gauge the performance of the different approximations, we consider a number of performance statistics. We tabulated the percentage of case - reorder point combinations for which the deviation between the approximation and the asymptotic simulation estimate was smaller than some threshold. This was repeated for different thresholds. We also

appr. method	difference $S_{\text{appr}} - S_{\text{opt}}$				
	< -1	-1	0	1	> 1
Optimal	0%	0%	100%	0%	0%
Dyn.-st.	0%	0%	91%	9%	0%
Average	1%	10%	83%	6%	0%
Bmrk	0%	0%	20%	58%	22%
"90%"	21%	26%	44%	10%	0%
"95%"	10%	23%	52%	15%	0%
"98%"	7%	16%	51%	26%	0%
"99%"	3%	10%	43%	43%	2%
"99,5%"	2%	7%	35%	51%	6%
"99,9%"	0%	2%	15%	53%	30%

Table 5: Deviations between the optimal reorder point and the reorder point that is deemed "optimal" based on the different approximations.

tabulated the maximum relative overestimate and the maximum relative underestimate of the different methods. Finally, we tabulated the average absolute relative deviation.

The table shows that the dynamic static waiting time approximation has excellent performance. It never over- or under- estimates the costs significantly. In some rare cases, it slightly overestimates the costs, but for almost all cases the deviation between approximation and asymptotic estimate is very small. The average relative deviation is only 0.7%. Note that target precision of the simulation was 0.1%, so the real precision of this approximation may even be higher.

The average waiting time approximation performs less well: it severely underestimates the downtime costs for some cases. The main reason is that the approximation ignores the fact that subsequent waiting times from the warehouse are dependent, which results in an underestimation of the length of downtime of multiple pieces of equipment in the same functionality group. Later on, we will see that in some cases, this has impact on the quality of the recommended reorder points. The main lesson is that, for redundant systems, the dependency of repair times through spare parts cannot be ignored without severe performance consequences. The dynamic static waiting time approximation does not have this problem. On the other hand, the average waiting time model has an advantage over the dynamic static waiting time approximation. The former is relatively easier to implement and to understand intuitively than the latter, and can for instance be easily implemented in a spreadsheet application.

Finally, the table shows that the benchmark method does not perform at all. We can conclude that detailed knowledge regarding redundancy is needed to obtain reasonable estimates of the long run average downtime costs.

6.3 Deviations from the true optimum

We are also interested in the quality of the reorder points that result from the different approximations. Note that a good performance regarding the precision of the approximations need not yield high quality reorder points. Conversely, the quality of the reorder points does not necessarily suffer from a bad approximation of the cost. Therefore, we make a separate assessment of the deviations of the approximations from the real optimum. We base the assessment of the quality of the reorder points on the cases for which the simula-

appr.	sum of costs over all cases ($\times 10^3 k\$/\text{yr}$)			Relative cost deviation				
method	holding	downtime	total	< 0.05	< 0.5	< 1	< 5	average
Optimal	1.73 (100%)	18.30(100%)	20.03(100%)	100%	100%	100%	100%	0%
Dyn.-st.	1.78 (102.7%)	18.27 (99.8%)	20.05 (100.1%)	100%	100%	100%	100%	0.05%
Average	1.65 (95.5%)	18.53(101.3%)	20.18(100.8%)	92.4%	99.2%	100%	100%	2.1%
Bmrk	2.65(152.8%)	18.04(98.6%)	20.69(103.3%)	59.8%	91.3%	96.5%	98.6%	38%
"90%"	1.53(88.3%)	21.66(118.4%)	23.19(115.8%)	56.5%	78.3%	86.1%	96.2%	82%
"94%"	1.82(105.0%)	18.84(103.0%)	20.66(103.2%)	70.1%	90.2%	94.0%	98.6%	40%
"98%"	2.06(119.0%)	18.49(101.0%)	20.55(102.6%)	70.4%	90.5%	95.4%	98.6%	38%
"99%"	2.37(136.8%)	18.13(99.1%)	20.50(102.4%)	69.6%	91.8%	96.2%	98.6%	37%
"99,5%"	2.55(147.0%)	18.08(98.8%)	20.62(103.0%)	65.2%	90.5%	95.7%	98.4%	51%
"99,9%"	3.08(177.7%)	18.01(98.4%)	21.09(105.3%)	51.9%	86.4%	92.9%	98.1%	74%

Table 6: The performance of the different heuristics with respect to the cost of using the approximately optimal reorder points.

tion attained the precision target. This means that 368 cases were taken into account, and 136 cases were omitted from the statistics. Again, while some cases were not taken into account, the considered cases still represent a wide range of systems, giving the statistics strong predictive power for other instances of the model. In addition to the approximations that were considered in Section 6.2, we now also consider the fill rate target traditional methods described in Section 4.4.

Let us first examine the deviations of the approximately optimal reorder points from the true optimum. The results are shown in Table 5. The dynamic static waiting time approximation again performs well, finding the optimal reorder point in 91% of the cases. For the other cases, it stocks one more than the optimum. We will see later that this has only relatively small cost consequences.

The average waiting time approximation performs less well. It finds the optimum in only 83% of the cases. Furthermore, in 1% of the cases it under stocks significantly. We will see that this has significant cost consequences.

The benchmark method (see Section 4.4), being unaware of redundancy, is unable to find the true optimum for many cases. The same conclusion can be drawn when looking at the deviations for the reorder points that are based on setting a service level target. While the risk of over (under) stocking can be reduced by decreasing (increasing) the service level target, it is impossible to find a single target that fits well on all cases. Detailed redundancy information is thus indispensable to find suitable reorder points for all parts.

6.4 Cost impact

Finally, we examine the cost impact of using the approximately optimal reorder points instead of the truly optimal ones. To assure that we obtain a complete picture of the performance of the approximate models, we use two different ways to aggregate the statistics of individual cases. First of all, we consider the real costs (as determined using simulation) of implementing the approximately optimal reorder points for all cases. These costs can be interpreted as the costs of implementing a certain policy at a company, i.e. for a diverse range of cases. Also, we gather statistics regarding the number of cases for which the relative deviation between the optimal costs and the cost of the approximately optimal reorder point (both costs determined using simulation) exceeds some threshold. The results are shown in

Table 6. The two ways to aggregate the information give a somewhat different perspective.

Let us now first discuss the cost totals over all different cases when using the different approximations. The dynamic static waiting time approximation performs well. In comparison to the optimal policy, the cost increase is 0.1%. Later on, we will argue that a more accurate comparison is obtained when focusing more on the relative stock increase. The dynamic static waiting time approximation uses 2.7% more stock than the optimal solution, of which $\sim 65\%$ is offset because of reduced downtime costs.

The average waiting time approximation has a slightly worse performance. It stocks somewhat less than the optimal policy, but incurs severe additional downtime costs as a result. The deteriorated performance is closely related to the fact that the average waiting time approximation significantly underestimates the downtime costs for some cases. The additional downtime costs are $0.23 \times 10^3 k\$/\mathrm{yr}$. This is about 14% of the total stock costs. The decrease in stock cost is only 4.5% of the total stock costs, resulting in a nett cost increase of 9.5% of the total stock cost, or 0.8% of the total cost.

Let us now discuss the other approximations. The benchmark method severely overstocks, but realizes some additional availability. However, this additional availability is not realized in a cost effective manner: the overall costs increase by 3.3%. Even though this might seem a small effect, comparing the cost increase with the total cost does not give an accurate picture. In an organization, the downtime costs are often hidden, since they translate to production that was not made. Moreover, a large fraction of the downtime costs are unavoidable, because even without any stock outs, downtime costs are still significant because of the remaining repair times. A more accurate picture of the impact on the company performance is obtained by concentrating on the holding cost (which are linear in the stock value). The benchmark method uses 52.8% more stock than the optimal stock quantity. Only $\sim 28\%$ of this 52.8% is offset by the increased availability. This means that the benchmark method leaves huge opportunities for improvement.

Note that in order to apply the benchmark method we do not need detailed information regarding the redundancy. A relevant practical question might thus be: will the costs of collecting this data be offset by the value of being able to make informed decisions (i.e. using the dynamic static waiting time approximation). The answer to this question is company -and case- specific, but we have shown that the potential for improvement is significant.

Let us now discuss the results for the fill rate target traditional methods, described in Section 4.4. The results for the use of the fill rate target show a significant cost increase with respect to the optimal reorder points. These results depend on the specific service level target that is set. When the service level target is too low (90%), availability is too low resulting in costly downtime. The cost increase of this is $3.36 \times 10^3 k\$/\mathrm{yr}$, which is more than two times the total optimal annual holding costs. Only a small fraction of this cost increase is offset by reduced holding costs. Conversely, by setting a very high service level target the holding cost increase significantly. When using a service level target of 99.9%, the holding cost increase by 177%. Only a small percentage of this cost increase is offset by improved availability, so the nett result is a cost increase larger than the total optimal annual holding costs.

The results improve somewhat if an intermediate service level target is chosen. When using a service level target of 98%, the realized availability is comparable to the availability that is realized when using the optimal policy. But this availability is realized at holding costs that are 19% higher.

Finally, we consider relative deviations of the cost of the approximative solution when compared with the cost of the optimal solution. We determine both costs using simulation.

The results show again that the performance of the dynamic static waiting time approximation is excellent: the cost increase is always lower than 5%, and the average relative deviation is only 0.05%.

The average waiting time clearly performs worse: in 0.8% of the cases it proposes solutions that have more than 50% cost increase in comparison to the cost of the optimal reorder point. For most cases however, it performs reasonable: it has an average relative deviation of 2.1%.

For the benchmark method, the relative deviation is large in some cases. For each benchmark method, in more than 1% of the cases the relative deviation is higher than 5. This means that the costs when using the approximation are more than 6 times as high as the cost of the optimal solution.

In conclusion, we have shown that the performance of the static dynamic waiting time approximation is good, leading to solutions that are optimal or close to optimal. We have identified some performance issues of the average waiting time approximation. Using two traditional methods, we have shown that ignoring the redundancy leads to significant cost increases, also when considering the total costs for a lot of cases. Finally, we have shown that for some cases, using methods that ignore the redundancy leads to solutions with a large cost deviation from the optimal solution.

7   Conclusions

In this study we have argued that in order to apply inventory models, we need a well-defined method to estimate shortage costs. We have argued that data coming from an RCM study can be a valuable source of information for this purpose. However, different downtime costs for different pieces of equipment and redundancy complicate the relation between shortage costs and the outcomes of the RCM study. In order to resolve this, we argued that an inventory model was needed capable of using the data from the RCM study. The model should be kept simple, because a very complex model would require too much effort on data collection to be applied. We developed such a model. The redundancy was modelled using functional groups, viz. groups of equipment for which the downtime costs depend on the number of pieces of equipment working within the group.

We developed approximate methods to determine the downtime costs from the model. We have shown how to find the reorder point using the methods. We also assessed the quality of the methods. We have shown that the dynamic static waiting time model has excellent performance. The average waiting time approximation has a somewhat degraded performance, but is more easy to implement.

Using two benchmark methods that mimic the approaches that are often applied in practice, we have shown that using detailed redundancy information can significantly improve the stocking decision.

Remark

The detailed numerical results including the results per case are available from the first author on request.

Acknowledgements

References

[1] W. D. Rustenburg, G. J. van Houtum, W. H. M. Zijm, Spare parts management at complex technology-based organizations: an agenda for research, International journal of production economics 71 (2001) 177–193.

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

[3] K. S. de Smidt-Destombes, M. C. van der Heijden, A. van Harten, On the availability of a k-out-of-N system given limited spares and repair capacity under a condition based maintenance strategy, Reliability Engineering & System Safety 83 (2004) 287 – 300.

[4] K. S. de Smidt-Destombes, M. C. van der Heijden, A. van Harten, On the interaction between maintenance, spare part inventories and repair capacity for a k-out-of-n system with wear-out, European Journal of Operational Research 174 (2006) 182 – 200.

[5] K. S. de Smidt-Destombes, M. C. vander Heijden, A. van Harten, Availability of k-out-of-N systems under block replacement sharing limited spares and repair capacity, International Journal of Production Economics 107 (2007) 404 – 421.

[6] K. S. de Smidt-Destombes, M. C. van der Heijden, A. van Harten, Joint optimisation of spare part inventory, maintenance frequency and repair capacity for k-out-of-n systems, International Journal of Production Economics 118 (2009) 260 – 268.

[7] S. R. Chakravarthy, A. Gómez-Corral, The influence of delivery times on repairable k-out-of-n systems with spares, Applied Mathematical Modelling 33 (2009) 23682387.

[8] R. Dekker, R. Plasmeijer, On the use of equipment criticality in maintenance optimization and spare parts inventory control, in: C. Guedes Soares (Ed.), Advances in safety & reliability, Vol. 3, ESRA, Pergamon press, Oxford, England, 1997, pp. 1709–1718.

[9] W. van Jaarsveld, R. Dekker, Integrating reliability centered maintenance and spare parts stock control, in: R. Briš, C. Guedes Soares, S. Martorell (Eds.), Reliability, risk, and safety: theory and application, ESRA, Taylor and Francis group, London, England, 1997, pp. 575–582.

[10] J. Moubray, Reliability-centered Maintenance, Butterworth Heinemann, Oxford, 1991.

[11] S. Axsäter, Inventory control, 2nd Edition, Springer, 2006.

[12] M. Kleijn, R. Dekker, An overview of inventory systems with several demand classes, in: M. Grazia Speranza, P. Stahly (Eds.), New trends in distribution logistics, Lecture notes in economics and mathematical systems 480, Springer Verlag, Berlin, Germany, 1999, pp. 253–266.

[13] F. Baskett, K. Chandy, R. Muntz, F. Palacios, Open, closed, and mixed networks of queues with different classes of customers, Journal of the association for computing machinery 22 (1975) 248260.

[14] C. C. Sherbrooke, METRIC: A Multi-Echelon Technique for Recoverable Item Control, Operations Research 16 (1968) 122–141.

[15] M. Trimp, S. Sinnema, R. Dekker, R. Teunter, Optimise initial spare parts inventories: an analysis and improvement of an electronic decision tool, Tech. Rep. EI 2004-52, Econometric institute, Erasmus University Rotterdam (2004).