Available online at www.sciencedirect.com
SCIENCE DIRECT®
Reliability Engineering and System Safety 83 (2004) 287–300
RELIABILITY ENGINEERING & SYSTEM SAFETY
www.elsevier.com/locate/ress

On the availability of a *k*-out-of-*N* system given limited spares and repair capacity under a condition based maintenance strategy

Karin S. de Smidt-Destombes[a,*], Matthieu C. van der Heijden[b], Aart van Harten[b]

[a]*TNO Physics and Electronics Laboratory, P.O. Box 96864, The Hague 2509 JG, The Netherlands*
[b]*University of Twente, Faculty of Business, Public Administration and Technology, Enschede, The Netherlands*

Received 30 April 2003; accepted 4 October 2003

Abstract

This paper considers a *k*-out-of-*N* system with identical, repairable components. Maintenance is initiated when the number of failed components exceeds some critical level. After a possible set-up time, all failed components are replaced by spares. A multi-server repair shop repairs the failed components. The system availability depends on the spare part stock level, the maintenance policy and the repair capacity. We present a mathematical model supporting the trade-off between these three parameters. We present both an exact and an approximate approach to analyse our model. In some numerical experiments, we provide insight on the impact of repair capacity, number of spares and preventive maintenance policy on the availability.

Keywords: Availability; Maintenance; Spare parts; Repair capacity; *k*-out-of-*N* systems

1. Introduction

Today's technological systems, such as aircrafts, nuclear power plants, military installations and advanced medical equipment, are characterised by a high level of complexity. The requirements for the availability and reliability of such systems are very high. Preventive maintenance is necessary to meet these requirements. To this end, we need sufficient resources, such as maintenance capacity and spare parts. Components that are very expensive are usually repaired after failure, if technically possible. The maintenance capacity is defined as the number of repairmen or repair facilities. This capacity is needed for both component repair and replacement. Throughout this paper we will also use the terms 'repair capacity' or just 'capacity' which are synonym for maintenance capacity.

So we have a trade-off between spare part inventories on the one hand and repair capacity on the other. Increasing the repair capacity implies a shorter throughput time for the spare parts in the repair process, and therefore less spares are needed. The other way around, increasing the number of repairable spare parts creates a buffer against uncertainties in the failure and repair process, which allows for a higher repair shop utilisation, and thus less repair capacity. Also, there is a (weaker) link between maintenance policy and the required repair capacity and number of spare parts, because more frequent maintenance will probably lead to more frequent repair jobs with less work content, and therefore to a more stable work load for the repair shop.

In this paper, we will examine the interaction between preventive maintenance policy, spare part inventories and repair capacity for a *k*-out-of-*N* system with exponentially distributed component life times and repair times. That is, a system consists of *N* identical components of which at least *k* components are needed for the system to perform its functions. All *N* components are subject to failure and have the same failure rate. This is called *hot standby* redundancy. As a variant of this model, we will also consider cold standby redundancy, where standby components cannot fail. In the latter of the two cases, we assume that the time to switch a component from standby state into operation is negligible. We consider situations with high set-up costs for maintenance, so that replacement of components is only initiated if a minimum number of components $m \le N - k$ has failed. A certain (deterministic) lead-time $L > 0$ is allowed between maintenance initiation and the start of

* Corresponding author. Tel.: +31-70-374-01-13; fax: +31-70-374-06-42.
E-mail address: desmidt@fel.tno.nl (K.S. de Smidt-Destombes).

0951-8320/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2003.10.004

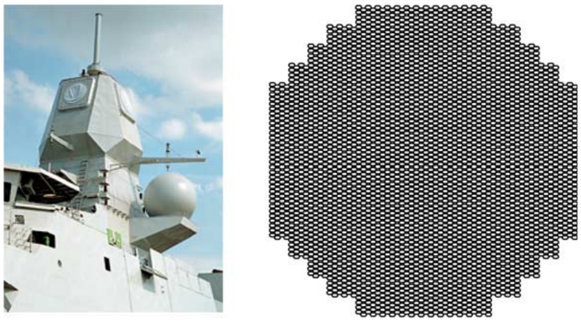

Fig. 1. The Active Phased Array Radar (APAR, left) consists of four 'faces', each having a large number of elements (right); a face can be modelled as a k-out-of-N system.

maintenance for the preparation of maintenance activities. We focus on calculating the availability of this system, given the number of repairable spare parts S and the repair capacity c.

Our study is motivated by several systems that we encountered at the Royal Netherlands Navy. An example is the Active Phased Array Radar (APAR), see Fig. 1. This radar has a cubical shape. On each of the four sides, it has a so-called *face*, consisting of thousands of transmit and receive elements. Each face covers a quarter of a circle, and together they cover the whole space around the ship of which it is a part. The elements on a face are identical and are partly redundant. A certain percentage of the total number of elements per face is allowed to fail, without losing the function of the specific radar face. Say that this percentage is 10% and that the total number of elements is 3000, then we have a 2700-out-of-3000 system[1].

To maintain the radar, it has to be taken off the frigate, because repair and replacement of elements has to be done in a dust-free environment and because special equipment and skills of the personnel are required. So, the set-up costs for maintenance are high. Therefore, maintenance is performed periodically only and not upon each element failure. Also, some lead-time is needed to prepare for maintenance, for example for the planning of the activities in the repair shop and possibly for navigating the frigate to the dockyard.

Another example for which a similar trade-off applies is the Active Towed Array Sonar (ATAS) for searching mines and submarines. The ATAS consists of several tens of hydrophones, let us assume 64 pieces. Say that 10% failed components is acceptable for full operation, then we model the ATAS as a 58-out-of-64 system. A smaller example is the frigate communication system (say, a 6-out-of-8 system).

[1] Because of confidential data, we only provide fictitious (but representative) numbers in this paper.

2. Literature

Several authors have mentioned preventive maintenance policy, spare parts and repair capacity as necessary pillars in an overall maintenance concept (see e.g. Ref. [5]). Only few papers actually deal with quantitative models integrating these elements. The best-known repairable spare parts models are the models based on METRIC, see Ref. [18]. Although METRIC uses the assumption that the capacity to restore spare parts is infinite, extensions have been made to include finite repair capacities. Various approaches have been considered, such as applying closed queuing networks [9], open queuing networks [23], Markov chains [2] and replacing infinite capacity queues by appropriate finite capacity queues ([19,24]). For a recent overview of the further spare part management literature, we refer to Refs. [13,16].

Only few papers deal with the simultaneous optimisation of spare parts and repair capacity. Ebeling [6] proposes a single echelon multi-item model where each item has its own resource capacity. A more general trade-off between repair capacity and spare part inventories, where multiple items share the same repair capacity, is given in Ref. [20]. Their model is different from ours, because (1) they do not consider hot standby redundancy and (2) they consider Poisson arrivals of failed parts at the repair shop and not batch arrival caused by high set-ups, as in our case.

Some work is available on the integration of maintenance and spares. Armstrong and Atkins [3,4] consider a model where one spare component is ordered when the used one has reached a certain age. If the spare is delivered before the component fails, it is kept in inventory. If failure before the ordering moment occurs it is possible, against higher cost, to get a spare quicker. The authors determine the cost per cycle. [10–12] use simulation to analyse an age replacement preventive maintenance policy where (non-repairable) spare parts are ordered based on a continuous review (s, S) inventory policy. For such systems replenishment is made whenever the inventory position drops below the reorder point s. This type of inventory system uses a variable replenishment quantity, which is enough to raise the inventory position up to level S. In our case we deal with repairable parts without scrap (probability of no repair possible is zero), which results in a constant number of spares equal to S. Sarkar and Haque [17] also develop a simulation model, in their case covering a block replacement policy and a continuous review policy for spare parts. Regarding the special case of redundant systems, Smith and Dekker [21] analyse a 1-out-of-N system with cold standby redundancy where an age replacement policy is applied to the operating component. In fact, the $N-1$ standby items act as spare parts.

To the best of our knowledge, only two papers address a maintenance policy and finite repair capacity for k-out-of-N systems. Fawzi and Hawkes [7] consider the case of hot standby redundancy, where failed components are sent to repair immediately upon failure. A single repair facility handles both component repair and component replacement, where pre-emptive priority is given to replacement activities. Frostig and Levikson [8] give a method to calculate the availability for k-out-of-N systems with both cold and warm standby redundancy. Again, spares are sent to the repair facility immediately upon failure. They consider a single server repair facility for the k-out-of-N case and a multi-server repair facility for the special case of 1-out-of-N systems only. As new aspects compared to these two papers, we will address (1) bulk arrival of failed components at the repair shop (caused by lead-times and costs), (2) a multi-server repair shop, and (3) maintenance lead times. For this model, we will determine the operational availability of a system for a combination of maintenance policy, number of spare parts and repair capacity.

The structure of this article is as follows. First, we describe the basic model in Section 3. In Section 4 we give an exact algorithm to determine the availability of a k-out-of-N system, depending on the maintenance policy and the resources needed. We present some simple approximations in Section 5. Next, we give numerical examples, illustrating the trade-off between capacity, maintenance policy and spare parts in Section 6. Section 7 deals with some model variants, such as the inclusion of replacement times and cold standby redundancy. In Section 8, we present our conclusions and we give some directions for further research.

3. Model and approach

3.1. Model and assumptions

In this section, we describe the k-out-of-N system with hot standby redundancy and its maintenance process in more detail. At the start of a system uptime, all N components are as good as new. The failure process of each component is characterised by a negative exponential distribution with rate λ, where we assume that the component failure processes are mutually independent. The system functions properly as long as at most $N-k$ components have failed. To prevent system downtime, maintenance is initiated if $m \leq N-k$ components have failed. It seems reasonable to choose $m = N-k$ if the maintenance set-up costs are high, but a lower number may be chosen if some lead-time $L \geq 0$ is required between maintenance initiation and the actual start of maintenance activities (which is true for the navy defence systems that motivated our research). The system is assumed to be in use during this lead-time and it is therefore likely to degrade further.

The actual maintenance activities consist of replacing all failed components by spares. However, if insufficient spares are available in an as-good-as-new condition, the maintenance completion is delayed until sufficient failed ones have been repaired. We assume that the components have independent and identical exponentially distributed repair times with rate μ. The capacity for restoring components is limited and equal to c parallel channels. For the time being, we ignore the replacement time of the components after repair (see Section 7 for an extension in this direction). When all failed components are replaced, the system cycle starts over again. During the time until the next maintenance initiation (i.e. when m components have failed) plus the lead-time L, the same capacity c is available for restoring components (see Section 7 for a generalisation to different repair capacities during system maintenance time and non-maintenance time). It is not guaranteed that the repair capacity is always sufficient to repair the remaining spares during the system uptime, so the number of available spares when maintenance starts may be less than S.

Our analysis in the remainder of this paper is based on the following additional assumptions

1. The failure process of components continues during the maintenance set-up time L, even if more than $N-k$ components have failed; the reason is that the APAR radar is always able to make partial observations in that case, so that the system will not be shut down; we refer to Section 7 for relaxing this assumption.
2. During maintenance, all failed components are replaced by new components; if it would be optimal to replace less components (say restoring up to $N_1 < N$), we have in fact an k-out-of-N_1 system; then, we conclude that too many components have been included in the system design.

For an overview of notation, we refer to Appendix A.

3.2. Approach

In fact, we have two interrelated cycles, namely, a cycle for the k-out-of-N system (uptime and downtime) and a cycle for the component repair process, see Fig. 2.

The system cycle starts with all N components as good as new. After maintenance initiation and the set-up period L, a number of n components have failed ($m \leq n \leq N$). During maintenance, these n components are replaced. Then, the system is restored and the next cycle starts. Because the initial state at the start of each cycle is the same, this seems to be a renewal process. The spares cycle starts at the beginning of the maintenance period, just before the k-out-of-N system comes in for maintenance. Then, s spare parts are available ($0 \leq s \leq S$), while the remaining $S-s$ spares still have to be repaired. If sufficient spares are available ($s \geq n$), all failed components are replaced and the system is operational again without delay. Otherwise, the system is down during the time to repair the remaining $n-s$ components needed. After maintenance completion, the repair process continues until the end of the cycle, i.e. just before the next maintenance period starts.

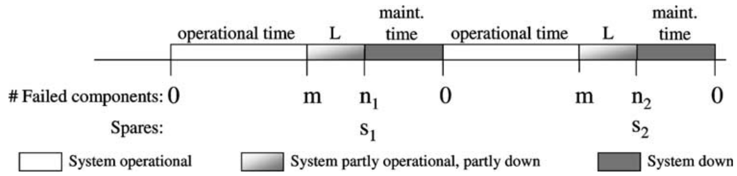

Fig. 2. Scheme of a system's cycle with the number of failed components (above) and the spares' cycle with the number of ready for use spares (beneath).

It is clear that the number of components at the start of a spares cycle depends on the number of components repaired during the cycle and the number of spares to be repaired at the start of the preceding cycle. Therefore, these cycles are dependent, and thus they do not constitute a renewal process. As a consequence, the combination of the two interrelated cycles is not a renewal process either. As a solution, we will derive the steady state distribution of the number of spares s at the start of a spares cycle. An exact steady state distribution provides us a way to an exact availability analysis.

The operational availability equals the expected uptime during a cycle (i.e. when at least k components are operational) divided by the expected cycle length. The expected uptime equals the expected time until maintenance initiation $E(T_m)$ plus the expected time during the set-up time L that at least k components are operational $E(U_m)$. So, we find

$$AV_{m,S,c} = \frac{E(T_m) + E(U_m)}{E(T_m) + L + E(D_{m,S,c})} \tag{1}$$

where $E(D_{m,S,c})$ is the expected maintenance time to restore the system to the new state. Eq. (1) implies that it is sufficient to find exact expressions for $E(T_m)$, $E(U_m)$ and $E(D_{m,S,c})$ as function of the three decision variables m, S and c.

4. An exact algorithm for the basic model

We first derive the expressions for $L = 0$, next we extend our analysis to a positive set-up time.

4.1. Zero set-up time ($L = 0$)

We only have to calculate $E(T_m)$ and $E(D_{m,S,c})$, as $E(U_m) = 0$. The operational time until maintenance initiation T_m can be derived by splitting this period in the time until the first component failure, the time between the first and the second failure, etc. The memoryless property of the exponential distribution gives us that the time between the ith and the $(i+1)$th failure is exponentially distributed with rate $(N-i)\lambda$. So, the expected time until the mth failure equals

$$E(T_m) = \sum_{i=0}^{m-1} \frac{1}{(N-i)\lambda} \tag{2}$$

To derive the expected maintenance duration $E(D_{m,S,c})$, we condition on the number of available spare parts s just before the system arrives for maintenance at the repair shop. Then, the system downtime equals the time for restoring the $m - s$ spares needed to repair the system

$$E(D_{m,s,c}) = \sum_{s=0}^{S} E[R_c(m-s, S-s+m|s)]\pi_{m,S,c}(s) \tag{3}$$

where $R_c(m-s, S-s+m|s)$ is the time to restore $m-s$ spares using c servers if $S-s+m$ components are waiting to be repaired, and $\pi_{m,S,c}(s)$ is the steady state probability of having s spares ready for use at the start of the maintenance period (just before the system arrives), given m, S and c.

Below, we will derive expressions for the two components involved in Eq. (3). We start with $E[R_c(i,j)]$, where we omit the conditioning variable s since it does not contain information and where we put $i = m-s$ and $j = S-s+m$ for simplicity. As obviously $E[R_c(i,j)] = 0$ if $i \le 0$, we focus on the case $i > 0$. Then, we can determine the expected maintenance period analogously to the derivation of $E[T_m]$ by splitting the period in the time until the first repair completion, the time between the first and the second repair completion, etc. We consider two situations, $j \le c$ and $j > c$. If $j \le c$, the time to restore the components is determined by the number of components to be restored j and not by the repair capacity c, so the mean time until the next repair completion equals $1/j\mu$. Otherwise, the repair capacity is the bottleneck, and the mean time until the next repair completion equals $1/c\mu$. In fact, we have the recursive relation

$$E[R_c(i,j)] = \frac{1}{\min\{j,c\}\mu} + E[R_c(i-1,j-1)] \tag{4}$$

We can elaborate this, finding the expression

$$ER_c(i,j) = \begin{cases} 0 & \text{if } i \le 0 \\ \sum_{h=0}^{i-1} \left\{ \frac{1}{(j-h)\mu} \right\} & \text{if } 0 < i \le j \le c \\ \frac{i}{c\mu} & \text{if } j > c \text{ and } i \le j-c \\ \frac{j-c}{c\mu} + \sum_{h=0}^{i-j+c-1} \left\{ \frac{1}{(c-h)\mu} \right\} & \text{if } j > c \text{ and } j-c < i \le j \end{cases} \tag{5}$$

We will determine the steady state probabilities $\pi_{m,S,c}(s)$ of having s spares ready for use at the start of the maintenance period (just before the system arrives) using a Markov chain. Because both failure and repair times are exponentially distributed, the transition probabilities solely depend on the state s at the beginning of a spares cycle. Each entry (i,j) of this matrix equals the probability $q_{i,j}$ that j spares are available at the start of a maintenance period while i spares were available at the start of the previous maintenance period ($i,j \in [0,\dots,s]$).

For computational efficiency, we first aggregate all states $s \le m$ in a single state M, so that the dimension of the Markov chain reduces from $S+1$ to $S-m+1$. The aggregation is useful, because we have insufficient spares available to repair the system immediately for all $s \le m$. Therefore, the number of spares to be repaired when the new system uptime starts equals S anyway, and so the probability of being in state σ at the start of the next cycle is the same for all $s \le m$. We will disaggregate the aggregate state M into states $s = 0,1,\dots,m$ later on. Note that we have $\pi_{m,S,c}(M) = 1$ as a special case if $S < m$, because we always have insufficient spares.

We calculate the transition probabilities $q_{i,j}$ by conditioning on the time to maintenance initiation $T_m = t$. Given that i spares are available just before a maintenance period starts and m spares are needed for repair, the number of spares to be repaired just after maintenance has started equals $S-i+m$. However, if insufficient spares are available ($i < m$), we have to wait until the number of spares available have increased to m, i.e. until the number of spares to be repaired has reduced to S. Hence, the number of spares to be repaired at the start of a system uptime equals $\min\{S, S-i+m\}$. This number has to be reduced to $S-j$ during the period $t+L$ to arrive in spares state j at the start of the next cycle. Therefore, we have

$$q_{i,j} = \int_0^\infty f_m(t) H_c(\min\{S, S-i+m\}, S-j, t)\mathrm{d}t \qquad (6)$$

where $f_m(t)$ is the density function of T_m and $H_c(a,b,\tau)$ is the probability that the number of failed spares will reduce from a to b during τ, i.e. exactly $a-b$ out of a spares will be repaired during τ with c servers. As $j = M$ represents the aggregate state $0,\dots,m$, $H_c(a, S-M, \tau)$ equals the probability that at most $a-S+m$ out of a spares will be repaired during τ. Because the number of component failures during t has a binominal distribution with parameters N and $p = 1-\mathrm{e}^{-\lambda t}$, we can derive that the density function $f_m(t)$ can be written as

$$f_m(t) = \binom{N}{m-1}(N-(m-1))\lambda \mathrm{e}^{-(N-(m-1))\lambda t}(1-\mathrm{e}^{-\lambda t})^{m-1} \qquad (7)$$

Regarding $H_c(a,b,\tau)$, we first note that only a positive number of components can be restored during τ, so that $H_c(a,b,\tau) = 0$ if $b > a$. If $a = b$, no components have been restored during τ. As the repair rate equals $\min\{b,c\}\mu$, we have that $H_c(b,b,t) = \mathrm{e}^{-\min\{b,c\}\mu t}$. For $b < a$. we distinguish two cases: $a \le c$ (all failed components are being repaired immediately) and $a > c$ (c repairs started initially). In the first case, the number of failed items remaining after a period t is binomially distributed with parameters a and $p = \mathrm{e}^{-\mu t}$. For the derivation of $H_c(a,b,\tau)$ if $b < a$ and $a > c$, we refer to Appendix B. Altogether, we find that

$$H_c(a,b,t) = \begin{cases} 0 & a<b \text{ or } a<0 \text{ or } b<0 \\ \mathrm{e}^{-\min\{b,c\}\mu t} & a=b \\ \binom{a}{b}\mathrm{e}^{-b\mu t}(1-\mathrm{e}^{-\mu t})^{(a-b)} & 0\le b\le a\le c \\ \displaystyle\sum_{g=0}^{c-b-1}\left[\binom{c}{b}\binom{c-b}{g}(-1)^g\left(\left(\frac{c}{(c-b-g)}\right)^{a-c} \times\left((\mathrm{e}^{-\mu(b+g)t}-\mathrm{e}^{-c\mu t}) - \sum_{h=1}^{a-c-1}\frac{c^{a-c}}{(c-b-g)^h} \times\frac{(\mu t)^{a-c-h}}{(a-c-h)!}\mathrm{e}^{-c\mu t}\right)\right] + \binom{c}{\min\{c,b\}} \times\frac{(-1)^{c-\min\{c,b\}}(c\mu t)^{a-\max\{c,b\}}}{(a-\max\{c,b\})!}e^{-c\mu t} & 0\le b\le a \text{ and } a>c \end{cases} \qquad (8)$$

Using Eqs. (7) and (8), we can find an explicit (but complicated) expression for the transition probabilities q_{ij} as defined by Eq. (6).

Next, we have to derive the steady state probabilities $\pi_{m,S,c}(i)$ for the states $0 \le i \le m$. We can use the following set of equations to derive these probabilities from the steady state probabilities $\pi_{m,S,c}(i)$, $m+1 \le i \le S$ and $\pi_{m,S,c}(M)$ for the aggregate state representing the states $0 \le i \le m$

$$\pi_{m,S,c}(i) = \pi_{m,S,c}(M)q_{M,i} + \sum_{j=m+1}^{S}\pi_{m,S,c}(j)q_{j,i} \qquad (9)$$

$0 \le i \le m$

For the transition probabilities q_{Mi}, we use the fact that S spares have to be repaired at the start of a system uptime if the spares state at the start of the cycle was $s \leq m$, no matter what the exact value of s was

$$q_{M,j} = \int_0^\infty f_m(t) H_c(S, S-j, t) \mathrm{d}t \tag{10}$$

Note that, as usual in Markov chains, we have a dependent system of equations, which we can solve by replacing one arbitrary equation by the condition that the sum of the entries of the vector $\pi_{m,S,c}(i)$ adds up to one. We can solve this system of equations using any standard numerical procedure, see e.g. Ref. [15].

Combining all stationary probabilities $\pi_{m,S,c}(s)$ with Eq. (5) we find $E(D_{m,S,c})$ from Eq. (3).

4.2. Positive set-up time ($L > 0$)

To solve the case $L > 0$, we extend our expressions. There are three consequences of a positive set-up time. First, we need the expected system uptime during maintenance set-up time $E(U_m)$, see Eq. (1), as the system fails if more than $(N-k-m)$ components fail during L. Secondly, the number of failed components in the system upon arrival at the repair shop is uncertain, because we have an additional number of component failures during L. Thirdly, the repair shop has more time to restore spares.

As the set-up time does not affect the expected operational time until maintenance initiation T_m, we can still use Eq. (2). The expected uptime during L depends on maintenance policy m. As the number of component failures during $t\,(0 \leq t \leq L)$ has a binomial distribution with parameters $N-m$ and $\mathrm{e}^{-\lambda t}$, the probability that the uptime exceeds t equals the probability that the number of failures during t is at most $N-m-k$. From this observation, we can derive that

$$E(U_m) = \sum_{i=0}^{N-m-k} \sum_{j=0}^{i} \binom{N-m}{N-m-i} \binom{i}{j} \times (-1)^j \frac{1-\mathrm{e}^{-(N-m-i+j)\lambda L}}{(N-m-i+j)\lambda} \tag{11}$$

For the expected maintenance duration $E(D_{m,S,c})$, we extend Eq. (3) by conditioning on the number of failed components in the system n as well. Then, the expected system downtime equals the time needed to restore the $n-s$ spares that are needed to repair the system

$$E(D_{m,s,c}) = \sum_{s=0}^{S} \sum_{n=m}^{N} E[R_c(n-s, S-s+n|n,s)] P_m(n) \pi_{m,S,c}(s) \tag{12}$$

where $P_m(n)$ is the probability that n components have failed at the start of system maintenance, given initiation upon failure of the mth component. This is the probability that $n-m$ components failed during the set-up time L. As the number of failures is binomially distributed with parameters $N-m$ and $1-\mathrm{e}^{-\lambda L}$, we find

$$P_m(n) = \binom{N-m}{n-m} \mathrm{e}^{-(N-n)\lambda L} (1-\mathrm{e}^{-\lambda L})^{(n-m)} \tag{13}$$

As the expression for $E[R_c(i,j)]$ remains identical to Eq. (5), we only have to modify the derivation of the steady state probabilities $\pi_{m,S,c}(i)$. To this end, we have to modify the transition probabilities q_{ij}, because we have to condition on both the time to maintenance initiation T_m and the number of component failures during the set-up time L

$$q_{ij} = \sum_{n=m}^{N} \left\{ P_m(n) \int_{t=0}^{\infty} f_m(t) H_c(\min\{S, S-i+n\}, S-j, t+L) \mathrm{d}t \right\} \tag{14}$$

We refer to Appendix C for an explicit expression of the transition probabilities q_{ij} that can be derived from Eq. (14), using Eqs. (7), (8) and (13). Given these modified transition probabilities, the approach remains the same. First, we aggregate all states $s \leq m$ to a single state M, then we solve the reduced Markov chain and finally we derive the state probabilities for $s \leq m$ from Eq. (9).

5. Approximation

Deriving the exact system availability given the decision variables S, c and m is not simple, because the expressions for the expected maintenance duration $E(D_{m,S,c})$ are complex. Therefore, we present an approximation in this section. We reduce the complexity by calculating the first two moments of the key stochastic variables involved rather than calculating the complete distribution. This approximation is based on the empirical finding that many stochastic systems are not very sensitive to the higher moments of the underlying probability distribution functions; see e.g. Ref. [22].

The expected maintenance time $E(D_{m,S,c})$ depends on the number of available spares just before the system arrives for maintenance, denoted by $B_{m,S,c}$ (having probability distribution $\pi_{m,S,c}(i)$). Let us further define A_m as the number of component failures during L (having a binominal distribution with parameters $N-m$ and $\mathrm{e}^{-\lambda L}$). Then, the number of components to be repaired during system maintenance equals $[m + A_m - B_{m,S,c}]^+$, where we denote $X^+ = Max\{X, 0\}$ for any variable X. If we assume that this number of components exceeds the number of parallel repair channels, we can approximate Eq. (5) as

$$ER_c(i,j) \approx \frac{i}{c\mu} \tag{15}$$

As a consequence, we can rewrite Eq. (12) as

$$E(D_{m,s,c}) \approx \frac{E\{[m + A_m - B_{m,S,c}]^+\}}{c\mu} \tag{16}$$

Now the idea is to use a two-moment approximation for the random variables A_m and $B_{m,S,c}$. That is, we calculate their first two moments and fit an appropriate distribution, such that the expected maintenance time $E(D_{m,S,c})$ can easily be approximated. We may approximate the distributions of A_m and $B_{m,S,c}$ by some discrete distributions or, more conveniently, by some continuous distributions if their mean is not too small (which is valid for large systems like the APAR). For continuous distributions, we may use Normal distributions or Erlang mixtures (cf. [22]). Normal distributions are more convenient, because the difference of two normally distributed random variables, $A_m - B_{m,S,c}$, is again normally distributed. Note that A_m has a binominal distribution, that converges to a normal distribution indeed if $N - m \to \infty$. For small numbers of components, a continuous approximation may be inaccurate. Then, Adan et al. [1] provide a method to fit a convenient discrete distribution to the first two moments of any discrete random variable on $\mathbf{Z}^+$. Depending on the mean and variance, a choice is made between a Poisson distribution and mixtures of binominal, negative binominal or geometric distributions.

To apply a moment approximation, we need to find the first two moments of A_m and $B_{m,S,c}$. The number of component failures during L, A_m, is binomially distributed, so that we have

$$E[A_m] = (N - m)[1 - e^{-\lambda L}], \tag{17}$$

$$\text{Var}[A_m] = (N - m)[1 - e^{-\lambda L}]e^{-\lambda L} \tag{18}$$

For the derivation of the first two moments of $B_{m,S,c}$, we use a stochastic equation, thereby avoiding the analysis of the Markov chain (9). As the demand for spares equals $m + A_m$, the number of spares available just before the next system uptime starts equals $[B_{m,S,c} - m - A_m]^+$. Let us define $Z_c(T_m + L)$ as the number of spares that can be repaired before the start of next maintenance period using c servers. Taking into account $Z_c(T_m + L)$, and the maximum number of spares that can be ready-for-use S, we find the following recursive relation

$$B_{m,S,c} = \text{Min}\{[B_{m,S,c} - m - A_m]^+ + Z_c(T_m + L), S\} \tag{19}$$

Unfortunately, $[B_{m,S,c} - m - A_m]^+$ and $Z_c(T_m + L)$ are mutually dependent. Therefore, we propose to approximate $Z_c(T_m + L)$ by $\tilde{Z}_c(T_m + L)$, being the number of spares that can be repaired before the start of next maintenance period using c servers if the number of items to be repaired is infinite. Then, we achieve that (1) $[B_{m,S,c} - m - A_m]^+$ and $Z_c(T_m + L)$ are mutually dependent, and (2) the moments of $\tilde{Z}_c(T_m + L)$ are easy to calculate (to be discussed below).

Now we can approximate the first two moments of $B_{m,S,c}$ applying the moment iteration approach that de Kok [14] introduced to analyse the $G/G/1$ queue. That is, given an initial estimate for the first two moments of $B_{m,S,c}$, we fit a simple (discrete or continuous) probability distribution function to the random variables $B_{m,S,c}$ and A_m. Based on these approximate distributions, we calculate the first two moments of $[B_{m,S,c} - m - A_m]^+$. This is straightforward if we use normal approximations, but more cumbersome for discrete approximations, given the diversity of specific distributions that we use. We solved the latter by brute force, i.e. by calculating the first two moments for each possible value of A_m, cutting the series off when the probability density has faded.

Next, we calculate the first two moments of $[B_{m,S,c} - m - A_m]^+ + \tilde{Z}_c(T_m + L)$. Then, again, we fit a (discrete or continuous) distribution to these first two moments and we calculate new approximations for the first two moments of $B_{m,S,c}$ from Eq. (19). We repeat these calculations until our approximations for the first two moments of $B_{m,S,c}$ converge. Although convergence is theoretically not guaranteed, application of this method has not led to convergence problems until now (de Kok, 1994).

To apply the moment-iteration approach to the recursive equation (19), we need the first two moments of $\tilde{Z}_c(T_m + L)$. We find these by conditioning on T_m. First, we note that we can find that the variance of the time until maintenance initiation, similarly to Eq. (2)

$$\text{Var}[T_m] = \sum_{i=0}^{m-1} \frac{1}{(N - i)^2\lambda^2} \tag{20}$$

The number of components that can be repaired during a period with length t is approximately Poisson distributed with mean $c\mu t$, provided that the workload of the repair shop is sufficiently high initially (and it is exact for $c = 1$). By conditioning on the length of the time until maintenance initiation T_m, we find that the mean and variance of $\tilde{Z}_c(T_m + L)$ equal

$$E[\tilde{Z}(T_m + L)] = c\mu\left\{L + \sum_{i=0}^{m-1} \frac{1}{(N - i)\lambda}\right\} \tag{21}$$

$$\text{Var}[\tilde{Z}_c(T_m + L)] = c\mu\left\{L + \sum_{i=0}^{m-1} \frac{1}{(N - i)\lambda}\right\} + (c\mu)^2 \sum_{i=0}^{m-1} \frac{1}{(N - i)^2\lambda^2} \tag{22}$$

Now we can approximate the expected maintenance time $E(D_{m,S,c})$using Eq. (16), because it holds that $E\{[m+A_m - B_{m,S,c}]^+\} = m + E[A_m] - E[B_{m,S,c}] + E\{[B_{m,S,c} - m - A_m]^+\}$ and $E\{[B_{m,S,c} - m - A_m]^+\}$ has been evaluated in the recursion (19).

Note that we can use a simple approximation for $E[U_m]$ using a moment-approach as well. To this end, we calculate the first two moments of the time to failure of a k-out-of-$(N - m)$ system from Eqs. (2) and (20), and we fit an Erlang-mixture on these two moments (cf. [22]). Let us denote the approximating distribution by T^*. Then we can easily calculate $E[U_m] \approx E[\min\{T^*, L\}]$. For a pure Erlang distribution, we have

$$E[\min\{T^*, L\}] = \frac{r}{\lambda}\left\{1 - \sum_{i=0}^{r} \frac{(\lambda L)^i e^{-\lambda L}}{i!}\right\} + L \sum_{i=0}^{r-1} \frac{(\lambda L)^i e^{-\lambda L}}{i!} \quad (23)$$

Finally, we note the moment iteration approach is very simple if we may assume normally distributed random variables, because (1) it is trivial to fit a normal distribution to the first two moments of a random variable, and (2) sums and differences of normally distributed random variables are normally distributed again. This seems a reasonable approach for large systems as the APAR.

6. Numerical experiments

We implemented the exact algorithm from Section 4 and the approximate algorithm from Section 5 (both the discrete and the continuous variant). During preliminary numerical tests, we found that our exact method works well for small and reasonable large systems, up until about 100 components. However, for very large numbers of components (say >100), we encountered numerical problems when calculating the transition probabilities from Eqs. (C.1) and (C.2). This is due to the extremely high binominal coefficients involved. Despite standard numerical tricks to reduce these computational problems (using recursive formulas and logarithms), stability problems remain for very large systems. Therefore, we have to use our approximate approach for such systems.

In this section, we first discuss numerical results for a moderately system size like the ATAS (58-out-of-64 system). We present trade-off figures between spare part inventory and repair capacity using our exact method. Running the same experiments for our approximate approach provides insight in the approximation accuracy. Next, we discuss numerical results for very large systems like the APAR (2700-out-of-3000) using our approximate method. We judge the accuracy of our approximation by comparison to results from discrete event simulation and we present trade-off figures.

6.1. Exact and approximate analysis of a 58-out-of-64 system (ATAS)

For a 58-out-of-64 system, maintenance can be initiated for some value of m between 1 and 6. We chose the set-up time L equal to 168 h ($=1$ week). We chose the time until component failure and component repair around eighteen months and one week, respectively, so $\lambda = 0.00008$ (failures/h) and $\mu = 0.006$ (repairs/h). We calculated the availability using our exact method for $c = 1, \ldots, 4$ and $S = 0, \ldots, 10$. The calculation time per case is less than 1 s.

In Fig. 3, we show the trade-off between the spare part inventory level S and the repair capacity c for this 58-out-of-64 system. We show the combinations of S and c yielding the same availability. For each point, we selected the maintenance initiation level m such, that the system availability is maximal (by enumeration over $m = 1, \ldots, 6$). We see that the only few spares are needed to compensate for less repair capacity if the target availability is low: both combinations $(S, c) = (1, 1)$ and (0,2) lead to an availability around 0.68. Considerably more spares are needed to compensate for repair capacity if the target availability is high: The combinations $(S, c) = (8, 1)$ and (3,2) are more or less equivalent for an availability around 0.95. Depending on specific cost parameters, a trade-off between spare part inventories and repair capacity can be made using Fig. 3.

To examine the impact of the maintenance control parameter m, we show in Fig. 4 the availability as function of m and S for a given repair capacity $c = 3$. If the criterion is to maximise availability, we see that the optimal value of m dependents on S: if $S = 0$, the availability increases

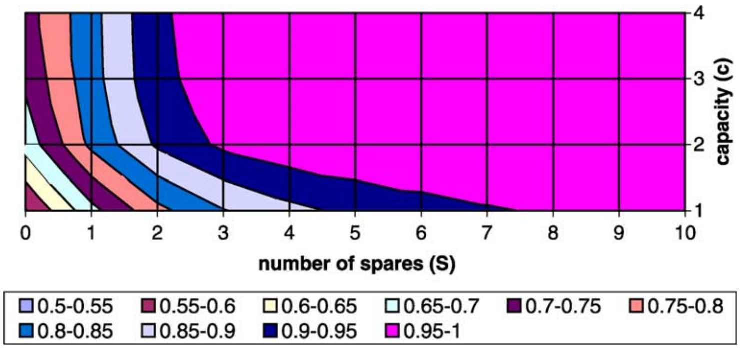

Fig. 3. The availability as function of the number of spares S and the repair capacity c, where the maintenance initiation level m has been chosen such that the availability is maximal.

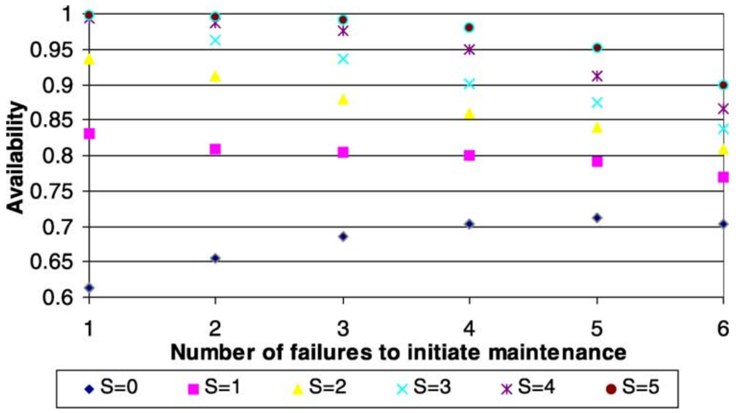

Fig. 4. The values of the availability for a combination of m, the number of failures until maintenance initiation for different values of the number of spares. The capacity is chosen equal to 3.

with m, whereas the availability decreases with m for $S \geq 1$. In the first case, the extra uptime gained from postponing maintenance initiation is larger than the extra downtime resulting from the component repairs. If we have spares available, however, it is better to initiate maintenance at the first failure. When the set-up costs for maintenance are high, this might not be the best value for m. Instead of having one spare and initiating maintenance at the first failure, we can also choose for three spares and initiate maintenance at the sixth failure. Both options give similar values for the availability but the cost involved can be very different.

We also used our approximate method to evaluate the same scenarios and we compared the results to the exact solution. We found that our approximations yield similar results. As can be expected, we find more accurate results using discrete probability distributions than using continuous (normal) probability distributions. The average error for the discrete and continuous approach is 0.28 and 0.87%, respectively, over 120 cases. The maximum error that we encountered is 4%, both for the discrete and the continuous approximation. The advantage of the continuous approximation is that it is much simpler and faster, because Eq. (20) is easier and faster to evaluate if all random variables are normally distributed. Very probably, the continuous approximation will be worse for smaller systems and better for larger systems.

6.2. Approximate analysis of a 2700-out-of-3000 system (APAR)

As a primary motivation for our research is the APAR radar, we analysed this system with the following fictitious parameters: $N = 3000$, $k = 2700$, $\lambda = 0.00008$, $\mu = 0.03$ and $L = 168$. In order to make trade-off figures, we calculated the availability for a large range of values for m (1…300), c (6…10) and S (5…200 with step size 5). Because we consider a very large system, we expect that the use of Normal distributions is probably as good as the use of discrete distributions. Surely, it is much faster. To check the accuracy of the approximation using Normal distributions, we simulated 25,000 cycles for a representative subset of 120 cases out of the parameter range above. We found that the deviation between approximate and simulated availability is

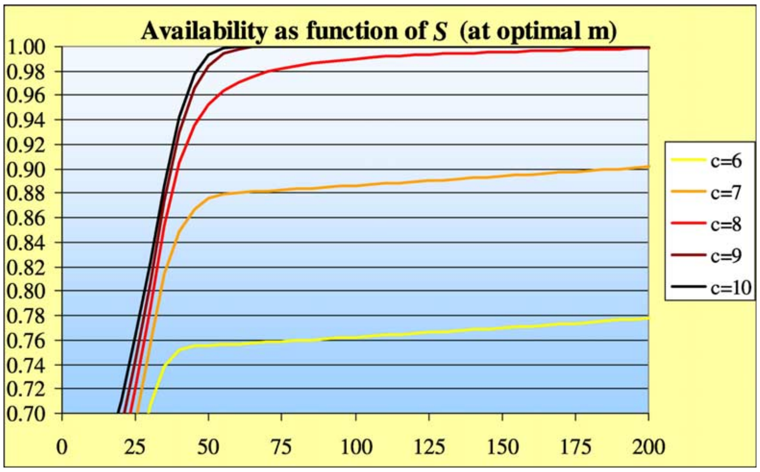

Fig. 5. The approximate availability as function of the spare part inventory level S for various repair shop capacities c (m chosen such that the availability is maximal).

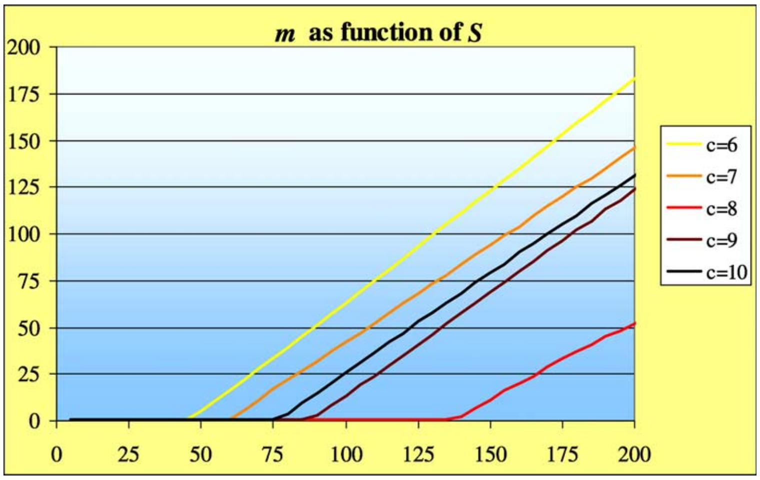

Fig. 6. Value of m for which the system availability is maximal.

0.15% on average with a maximum of 1.64%. The most serious approximation errors occur if $m = 1$. For more reasonable values of m (we further tested $m = 50$, 150, 25), the deviation between approximation and simulation is only 0.02% on average (and 0.25% maximum). Therefore, we conclude that it is safe to use normal distributions.

In Figs. 5 and 6, we show the main results from our numerical experiments. The first figure gives the approximate availability as function of the spare part inventory level S for various repair shop capacities c, where m has been chosen such that the approximate availability is maximal. The corresponding values of m are given in Fig. 6. We see that remarkably small values of m are optimal if we use the system availability as criterion, irrespective of costs. If the number of spare parts is somewhat small (which could occur if these spare parts are very expensive) and maintenance set-up costs are negligible, it is better to repair the system more frequently. After a certain spare part level, m increases almost linearly with the spare part stock level (i.e. the maintenance frequency decreases).

Figures like the two as shown in this section can be used to make a trade-off between spare part inventories and repair capacities if the relevant cost factors are known. For optimisation, enumeration is an option, because the approximation based on normal distributions is very fast (tens of thousands of problem instances can be evaluated within one minute CPU time on a Pentium III 700 MHz PC).

7. Model variants

In this section, we will discuss a some model extensions and variants, namely (1) the repair capacity during system uptime and maintenance time is different, (2) the component failure process stops if less than k components are available, (3) cold standby redundancy, and (4) account for component replacement times.

7.1. Sufficient repair capacity

We can simplify the expressions from Section 4 considerably if we assume that the repair capacity is sufficient to repair all spares during the time the system is not maintained, $T_m + L$. In that case, it holds that $\pi_{m,S,c}(S) = 1$ and $\pi_{m,S,c}(i) = 0$, $0 \leq i \leq S - 1$, and so Eq. (12) is simplified as

$$E(D_{m,s,c}) \approx \sum_{n=\text{Max}\{S,m\}}^{N} E[R_c(n - S, n|n)]P_m(n) \qquad (24)$$

Now we can evaluate Eq. (24) simply by substitution of Eqs. (5) and (13). A drawback of this approximation is that it does not facilitate a proper trade-off between maintenance policy, spare part inventory and repair capacity. Reducing the repair capacity may lead to a serious violation of our approximating assumption, so that our approximation will become very inaccurate.

7.2. Repair capacity during ($T_m + L$) differs from the capacity during maintenance time

When a system fails and the number of spares is insufficient, it is possible that additional repair capacity will be deployed. Suppose that the normal repair capacity (during $T_m + L$) equals c_1 and that the capacity during maintenance equals $c_2 > c_1$. We can easily incorporate this refinement by using repair capacity $c = c_1$ in Eq. (8), affecting the steady state probabilities $\pi_{m,S,c}(i)$, and repair capacity $c = c_2$ in Eq. (12), affecting the mean system maintenance time $E(D_{m,S,c})$.

7.3. System is shut down after more than Nk component failures

If the system shuts down when less than k components are available, the component failure process can stop before maintenance starts. The only expression that has to be modified in that case is the distribution of the number of failed items in the system when maintenance starts, $P_m(n)$, because we have an upper bound on the number of failed items. As a consequence, expression (13) remains valid for $m \le n \le N-k$, but the probability mass for all $n \ge N-k+1$ is concentrated in $N-k+1$

$$P_m(N-k+1) = \sum_{i=N-k+1-m}^{N-m} \binom{N-m}{i} \times e^{-(N-m-i)\lambda L}(1-e^{-\lambda L})^i \quad (25)$$

7.4. Cold standby redundancy

Let us assume that components cannot fail during standby status and that the system is shut down if less than k components are available. Then we have to modify the expressions regarding the failure process. As the mean time between two successive component failures in the k-out-of-N system equals $1/k\lambda$, the time until maintenance initiation has an Erlang-m distribution with scale parameter $k\lambda$, so we modify Eq. (7) and Eq. (2), respectively as

$$f_m(t) = \frac{(k\lambda)^m t^{m-1} e^{-k\lambda t}}{(m-1)!} \quad (26)$$

$$E(T_m) = \frac{m}{k\lambda} \quad (27)$$

The probability that n components have failed at the start of system maintenance $P_m(n)$ can easily be derived, as the number of component failures during the set-up time L is Poisson distributed, with all mass for $n \ge N-k+1$ being concentrated in $N-k+1$

$$P_m(n) = \begin{cases} \dfrac{(k\lambda L)^{n-m} e^{-k\lambda L}}{(n-m)!} & m \le n \le N-k \\ 1 - \displaystyle\sum_{i=0}^{N-k-m} \frac{(k\lambda L)^i e^{-k\lambda L}}{i!} & n = N-k+1 \end{cases} \quad (28)$$

To derive the mean system uptime during maintenance set-up $E(U_m)$, we use that the probability of this uptime exceeding t equals the probability that at most $(N-k-m)$ components fail until t. As this number of failures is Poisson distributed with mean $k\lambda$, we find

$$E(U_m) = \int_0^L \Pr\{U_m > t\} dt = \int_0^L \sum_{i=0}^{N-k-m} \frac{(k\lambda t)^i e^{-k\lambda t}}{i!} dt$$

Some algebra yields

$$E(U_m) = \frac{N-k-m+1}{k\lambda} - \frac{e^{-k\lambda L}}{k\lambda} \times \sum_{j=0}^{N-k-m} \frac{(N-k-m-j+1)(k\lambda L)^j}{j!} \quad (29)$$

We obtain an analytic expression for the transition probabilities q_{ij} by substituting the expressions above in Eq. (14).

7.5. Including component replacement times

Next to component repair, component replacement is a part of the maintenance activities. Let us assume that the time required for a single component replacement v is deterministic and that the same repair capacity is needed for component repair and replacement (otherwise the model extension is trivial). Component replacement occurs as soon as sufficient components have been repaired. Then, the system availability should be calculated as

$$A_{m,S,c} = \frac{E(T_m) + E(U_m)}{E(T_m) + L + E(D_{m,S,c}) + E[V_m]} \quad (30)$$

where V_m denotes the time needed for component replacement. If all repair capacity is used for component replacement, the fact that the number of failures during the set-up time L is binominally distributed (see Eq. (13)) leads us to

$$E[V_m] = v\left\lceil \frac{\{m + (N-m)(1-e^{-\lambda L})\}}{c} \right\rceil \quad (31)$$

where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. However, if only a single repair man is used for component replacement while the remaining capacity $(c-1)$ is used for repair, the steady state probabilities $\pi_{m,S,c}(i)$ should be modified as well, which influences $E(D_{m,S,c})$. For the transition probabilities q_{ij}, we have to take into account that $c-1$ servers are available to repair components during the replacement time V

$$q_{ij} = \sum_{n=m}^{N} P_m(n) \sum_{h=0}^{j} H_{m,S,c-1}(\min\{S, S-i+n\}, S-h, nv) \times \int_{t=0}^{\infty} f_m(t) H_{m,S,c}(S-h, S-j, t+L) dt \quad (32)$$

It should be possible to derive a closed form expression for q_{ij}, but it is clear that this is complex.

8. Conclusions and further research

In this paper, we presented both an exact and an approximate method to make a trade-off between spare part inventories, repair capacity and maintenance policy for a simple model. The exact method works very well for systems up to 100 components, for larger systems the approximate method can be used. If the number of components is high, as for the APAR radar, we recommend to use Normal distributions for convenience and to reduce computational effort.

Although we discussed various model extensions, it is clear that our model is just a first step towards the integration of spare part management and preventive maintenance optimisation. Practical situations tend to be far more complex than the simple model that we addressed here. Relevant issues include (a) multiple systems sharing the same repair capacity, (b) multiple failure modes (c) multiple item types that can be repaired by the same repair shop, and (d) component wear-out (i.e. non-exponential failure behaviour). A model taking into account these aspects has to be far more extensive. Our future research will deal with these issues. Below, we give some ideas how to proceed.

To include the aging of components, we could add one or more states representing the component quality (state 0 is as good as new state, state 1 is degraded and state 2 is failed). Our current model may be extended in this direction if we assume an exponential sojourn time in each state. An additional complication is that we have different component types sharing the same repair capacity (both failed and degraded components can be repaired). Then, we need to decide in which order the items have to be repaired. Also, the decision when to initiate maintenance activities may be determined by both the number of failed components and the number of degraded components. The extension of our models taking into account these aspects is the next step in our research activities.

Appendix A. Notation

c	repair capacity
k	the least number of components needed for a functional system
L	lead-time: the time from maintenance initiation until the start of maintenance activities
m	the number of failed components to initiate maintenance
N	the total number of components in the system
S	the total number of spares
λ	the failure rate of a single component
μ	the repair rate of a single component
A_m	the number of component failures during L, given the maintenance initiation level m
$B_{m,S,c}$	the number of ready for use spares just before the start of maintenance as a function of the maintenance initiation level m, the total number of spares S and the repair capacity c
$f_m(t)$	density function of the time duration t until maintenance initiation, given maintenance initiation level m
$H_c(a,b,t)$	probability to reduce the number of failed spares from a to b during time t, given repair capacity c
$P_m(n)$	probability that n system components are failed at the start of maintenance, given maintenance initiation level m
$q_{i,j}$	probability to start a spares cycle with j ready for use spares if the number of ready for use spares at the previous cycle was equal to i
$R_c(i,j)$	time needed to restore i spares of the total j spares to be restored with capacity c
$Z_c(t)$	number of spares that can be repaired before the next maintenance period, given repair capacity c
$\tilde{Z}_c(t)$	approximation of $Z_c(t)$, assuming the number of spares to be repaired as infinite
$\pi_{m,S,c}(s)$	the steady state probability of having s ready for use spares at the maintenance start as a function of the maintenance initiation level m, the total number of spares S and the repair capacity c
$AV_{m,S,c}$	the system availability, given the maintenance initiation level m, the total number of spares S and the repair capacity c
T_m	time until maintenance initiation given maintenance initiation level m
U_m	uptime during the lead-time L, given maintenance initiation level m
$D_{m,S,c}$	downtime caused by maintenance activities, given maintenance initiation level m, the total number of spares S and the repair capacity c

Appendix B. The derivation of $H_c(i,j,\tau)$

In this appendix, we derive an expression for $H_c(i,j,\tau)$, the probability that the number of failed spares will reduce from i to j during τ, if $i > c$. Then the number of spares to restore exceeds c, but only c spares can be repaired simultaneously. Let τ be the time at which the first repair is completed. In the remaining time $t-\tau$, $i-1-j$ out of $i-1$ failed components have to be restored. Hence,

$$H_c(i,j,t) = \int_{\tau=0}^{t} c\mu e^{-c\mu\tau} H_c(i-1,j,t-\tau)d\tau \qquad (B.1)$$

We distinguish two situations, $j < c$ and $j \geq c$. In the first situation, we start with $H_c(c+1,j,t)$:

$$
\begin{aligned}
H_c(c+1,j,t) &= \int_{\tau=0}^{t} c\mu e^{-c\mu\tau} H_c(c,j,t-\tau)\mathrm{d}\tau \\
&= \int_{\tau=0}^{t} c\mu e^{-c\mu\tau} \binom{c}{j} e^{-j\mu(t-\tau)}(1-e^{-\mu(t-\tau)})^{c-j}\mathrm{d}\tau \\
&= \sum_{h=0}^{c-j} \binom{c}{j}\binom{c-j}{h}(-1)^h c\mu e^{-(j+h)\mu t} \\
&\quad \times \int_{\tau=0}^{t} e^{-(c-j-h)\mu\tau}\mathrm{d}\tau \\
&= \sum_{h=0}^{c-j-1}\Bigg[\binom{c}{j}\binom{c-j}{h}(-1)^h \\
&\quad \times \frac{c\mu e^{-(j+h)\mu t}(1-e^{-(c-j-h)\mu t})}{(c-j-h)\mu}\Bigg] \\
&\quad + \binom{c}{j}(-1)^{c-j} c\mu t e^{-c\mu t} \\
&= \sum_{h=0}^{c-j-1}\left[\binom{c}{j}\binom{c-j}{h}(-1)^h \frac{c(e^{-(j+h)\mu t}-e^{-c\mu t})}{(c-j-h)}\right] \\
&\quad + \binom{c}{j}(-1)^{c-j} c\mu t e^{-c\mu t}
\end{aligned}
\qquad \text{(B.2)}
$$

In this way, we can calculate $H_c(i,j,t)$ recursively for $i=c+2$, $i=c+3$, etc. Repeating this, we find Eq. (8).

If $c \le j < i$ we use Eq. (10) and we start with $i = j+1$

$$
\begin{aligned}
H_c(j+1,j,t) &= \int_{\tau=0}^{t} c\mu e^{-c\mu\tau} H_c(j,j,t-\tau)\mathrm{d}\tau \\
&= \int_{\tau=0}^{t} c\mu e^{-c\mu\tau} e^{-c\mu(t-\tau)}\mathrm{d}\tau = c\mu t e^{-c\mu t}
\end{aligned}
$$

Again, we can calculate $H_c(i,j,t)$ recursively for $i = c+2$, $i = c+3$, etc. Then we find that

$$
H_c(i,j,t) = \frac{(c\mu t)^{i-j}}{(i-j)!} e^{-c\mu t}
$$

Appendix C. The transition probabilities q_{ij}

Regarding the transition probabilities q_{ij} as defined in Eq. (14), we can rewrite q_{ij} such that the integral is eliminated. We distinguish the case $i \le m$ and the case $i > m$. In the first case, Eq. (14) can be written as

$$
\begin{aligned}
q_{ij} &= \sum_{n=m}^{N}\left\{P_m(n)\int_{t=0}^{\infty} f_m(t) H_c(\min\{S,S-i+n\},S-j,t+L)\mathrm{d}t\right\} \\
&= \sum_{n=m}^{N}\{P_m(n)\}\int_{t=0}^{\infty} f_m(t) H_c(S,S-j,t+L)\mathrm{d}t \\
&= \int_{t=0}^{\infty} f_m(t) H_c(S,S-j,t+L)\mathrm{d}t
\end{aligned}
$$

Substituting $f_m(t)$ as defined in Eq. (7) and $H_c(S,S-j,t+L)$ as defined in Eq. (8), we find expression (B.1) if $j=0$ or $S \le c$ and expression (B.2) if $j>0$ and $S>c$.

$$
\begin{aligned}
&\binom{N}{m-1}(N-m+1)\lambda\binom{S}{S-j} \\
&\times \sum_{h=0}^{m-1}\sum_{g=0}^{j}\left\{\binom{j}{g}\frac{(-1)^{g+h} e^{-(\min\{c,S-j\}+g)\mu L}}{(N-m+1+h)\lambda+(\min\{c,S-j\}+g)\mu}\right\}
\end{aligned}
\qquad \text{(C.1)}
$$

$$
\begin{aligned}
&\binom{N}{m-1}(N-m+1)\lambda\Bigg[\binom{c}{\min\{c,S-j\}} \\
&\times\Bigg((c\mu)^{S-\max\{c,S-j\}}\sum_{h=0}^{m-1}\sum_{d=0}^{S-\max\{c,S-j\}}\Bigg\{\binom{m-1}{h} \\
&\times\frac{(-1)^{c-\min\{c,S-j\}+h} L^{S-\max\{c,S-j\}-d} e^{-c\mu L}}{(S-\max\{c,S-j\}-d)!((N-m+1+h)\lambda+c\mu)^{d+1}}\Bigg\} \\
&+\binom{c}{S-j}\sum_{g=0}^{c-S+j-1}\Bigg\{\binom{c-S+j}{g}\Bigg(\left(\frac{c}{c-S+j-g}\right)^{S-c}\sum_{h=0}^{m-1}\Bigg\{\binom{m-1}{h} \\
&\times\left(\frac{(-1)^{g+h} e^{-(S-j+g)\mu L}}{(N-m+1+h)\lambda+(S-j+g)\mu}-\frac{(-1)^{g+h} e^{-c\mu L}}{(N-m+1+h)\lambda+c\mu}\right)\Bigg\} \\
&+-\sum_{h=1}^{S-c-1}\Bigg\{\frac{c^{S-c}\mu^{S-c-h}}{(c-S+j-g)^h}\sum_{f=0}^{m-1}\sum_{d=0}^{S-c-h}\Bigg\{\binom{m-1}{f} \\
&\times\frac{(-1)^{f+g} L^{S-c-h-d} e^{-c\mu L}}{(S-c-h-d)!((N-m+1+f)\lambda+c\mu)^{d+1}}\Bigg\}\Bigg\}\Bigg)\Bigg\}\Bigg)\Bigg]
\end{aligned}
\qquad \text{(C.2)}
$$

In the second case, $i>m$, we split Eq. (14) into two parts

$$
\begin{aligned}
q_{ij} &= \sum_{n=m}^{N}\Bigg\{P_m(n)\int_{t=0}^{\infty} f_m(t) H_c(\min\{S,S-i+n\}, \\
&\qquad S-j,t+L)\mathrm{d}t\Bigg\} \\
&= \sum_{n=m}^{i-1}\left\{P_m(n)\int_{t=0}^{\infty} f_m(t) H_c(S-i+n,S-j,t+L)\mathrm{d}t\right\} \\
&\quad + \sum_{n=i}^{N}\left\{P_m(n)\right\}\int_{t=0}^{\infty} f_m(t) H_c(S,S-j,t+L)\mathrm{d}t
\end{aligned}
$$

For the second part we can use the expressions found in case 1. For the first part, we find

$$
\begin{aligned}
&\binom{N}{m-1}(N-m+1)\lambda\Bigg[\sum_{n=m}^{\min\{N,i-1,c-S+i\}}\Bigg\{P_m(n)\binom{S-i+n}{S-j} \\
&\quad\times\sum_{h=0}^{m-1}\sum_{g=0}^{n-i+j}\Bigg\{\binom{m-1}{h}\binom{n-i+j}{g} \\
&\quad\times\frac{(-1)^{g+h}\mathrm{e}^{-(S-j+g)\mu L}}{(N-m+1+h)\lambda+(S-j+g)\mu}\Bigg\}\Bigg\} \\
&\quad+\sum_{n=1+\min\{i-1,c-S+i\}}^{\min\{i-1,N\}}\Bigg\{P_m(n)\binom{c}{\min\{c,S-j\}} \\
&\quad\times(c\mu)^{S-i+n-\max\{c,S-j\}}\sum_{h=0}^{m-1}\sum_{d=0}^{S-i+n-\max\{c,S-j\}}\Bigg\{\binom{m-1}{h} \\
&\quad\times\frac{(-1)^{c-\min\{c,S-j\}+h}L^{S-i+n-\max\{c,S-j\}-d}\mathrm{e}^{-c\mu L}}{(S-i+n-\max\{c,S-j\}-d)!((N-m+1+h)\lambda+c\mu)^{d+1}}\Bigg\} \\
&\quad+P_m(n)\binom{c}{S-j}\sum_{g=0}^{c-S+j-1}\Bigg\{\binom{c-S+j}{g} \\
&\quad\Bigg(\Bigg(\frac{c}{c-S+j-g}\Bigg)^{S-i+n-c}\sum_{h=0}^{m-1}\Bigg\{\binom{m-1}{h} \\
&\quad\Bigg(\frac{(-1)^{g+h}\mathrm{e}^{-(S-j+g)\mu L}}{(N-m+1+h)\lambda+(S-j+g)\mu}-\frac{(-1)^{g+h}\mathrm{e}^{-c\mu L}}{(N-m+1+h)\lambda+c\mu}\Bigg)\Bigg\} \\
&\quad+-\sum_{h=1}^{S-i+n-c-1}\Bigg\{\frac{c^{S-i+n-c}\mu^{S-i+n-c-h}}{(c-S+j-g)^h} \\
&\quad\times\sum_{f=0}^{m-1}\sum_{d=0}^{S-i+n-c-h}\Bigg\{\binom{m-1}{f} \\
&\quad\times\frac{(-1)^{f+g}L^{S-i+n-c-h-d}\mathrm{e}^{-c\mu L}}{(S-i+n-c-h-d)!((N-m+1+f)\lambda+c\mu)^{d+1}}\Bigg\}\Bigg\}\Bigg)\Bigg\}\Bigg\}\Bigg\}\Bigg]
\end{aligned}
$$

References

[1] Adan IJBF, van Eenige MJA, Resing JAC. Fitting discrete distributions on the first two moments. Probab Engng Info Sci 1996;9:623–32.

[2] Albright SC, Gupta A. Steady-state approximation of a multi-echelon multi-indentured repairable-item inventory system with a single repair facility. Naval Res Logist 1993;40(4):479–93.

[3] Armstrong MJ, Atkins DR. Joint optimization of maintenance and inventory policies for a simple system. IIE Trans 1996;28:415–24.

[4] Armstrong MJ, Atkins DR. A note on joint optimization of maintenance and inventory. IIE Trans 1998;30:143–9.

[5] Dinesh Kumar U, Crocker J, Knezevic J, El-Haram M. Reliability, maintenance and logistic support: a life cycle approach. Norwell: Kluwer; 2000.

[6] Ebeling CE. Optimal stock levels and service channel allocations in a multi-item repairable asset inventory system. IIE Trans 1991;23: 115–20.

[7] Fawzi BB, Hawkes AG. Availability of an *R*-out-of-*N* system with spares and repairs. J Appl Probab 1991;28:397–408.

[8] Frostig E, Levikson B. On the availability of *R* out of *N* repairable systems. Naval Res Logist 2002;49(5):483–98.

[9] Gross D, Miller DR, Soland RM. A closed queueing network model for multi-echelon repairable item provisioning. IIE Trans 1983;15(4): 344–52.

[10] Kabir ABMZ, Al-Olayan AS. Joint optimization of age replacement and continuous review spare provisioning policy. Int J Oper Product Manage 1994;14(7):53–69.

[11] Kabir ABMZ, Al-Olayan AS. A stocking policy for spare part provisioning under age based preventive replacement. Eur J Oper Res 1996;90:171–81.

[12] Kabir ABMZ, Farrash SHA. Simulation of an integrated age replacement and spare provisioning policy using SLAM. Reliab Engng Syst Saf 1996;52(2):129–38.

[13] Kennedy WJ, Patterson JW, Fredendall LD. An overview of recent literature on spare parts inventories. Int J Product Econom 2002;76: 201–15.

[14] de Kok AG. A moment-iteration method for approximating the waiting time characteristics of the G/G/1 queue. Probab Engng Info Sci 1989;3:273–87.

[15] Press WH. Numerical recipes in C + 2002.

[16] Rustenburg WD, van Houtum GJ, Zijm WHM. Spare parts management at complex technology-based organizations: an agenda for research. Int J Product Econom 2001;71:177–93.

[17] Sarker R, Haque A. Optimization of maintenance and spare provisioning policy using simulation. Appl Math Model 2000;24: 751–60.

[18] Sherbrooke CC. Optimal inventory modeling of systems: multi-echelon techniques. New York: Wiley; 1992.

[19] Sleptchenko MC, van der Heijden MC, van Harten A. Effects of finite repair capacity in multi-echelon, multi-indenture service part supply systems. Int J Product Econom 2002;79:209–30.

[20] Sleptchenko A, van der Heijden MC, van van Harten A. Trade-off between inventory and repair capacity in spare part networks. J Oper Res Soc 2003;54(3):263–72.

[21] Smith MAJ, Dekker R. Preventive maintenance in a 1-out-of-*n* system: the uptime, downtime and costs. Eur J Oper Res 1997;99: 565–83.

[22] Tijms HC. Stochastic models: an algorithmic approach. Chichester: Willey; 1994.

[23] Zijm WHM, Avsar ZM. Capacitated two-indenture models for repairable item systems. Int J Product Econom 2003;81–82(C): 573–88.

[24] Diaz A, Fu MC. Models for multi-echelon repairable item inventory systems with limited repair capacity. European Journal of Operations Research 1997;97:480–92.