Available online at www.sciencedirect.com
ScienceDirect
Int. J. Production Economics 107 (2007) 404–421
international journal of production economics
www.elsevier.com/locate/ijpe

Availability of k-out-of-N systems under block replacement sharing limited spares and repair capacity

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

[a] *TNO Defence, Security and Safety, P.O. Box 96864, 2509 JG, The Hague, The Netherlands*
[b] *University of Twente, Enschede, The Netherlands*

Received 4 January 2005; accepted 25 August 2006
Available online 15 February 2007

Abstract

In this paper we consider an installed base of k-out-of-N systems, each consisting of identical, repairable components. A block replacement policy is used to maintain each system and all components are repaired by a single repair shop. System maintenance consists of replacing all failed and degraded components by spares. We focus on the downtime resulting from the lack of spare parts. The control variables that influence the system availability are the maintenance interval, the spare part inventory level and the repair capacity. We present two approximate methods to analyse the relation between these control variables and the system availability. Comparison with simulation results shows that we can generate nearly accurate approximations for the system availability using one of these models, depending on the system size. The average errors are found to be between 0.1% and 4.3%, compared to simulation. We found that the errors become smaller when the installed base increases and the number of system components becomes larger.

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

1. Introduction

Today's technological systems like aircraft, military or medical equipment are becoming more complex. At the same time, the users demand a very high availability which can be achieved in several ways. First, redundancy of critical components can be included in the system design. Second, the system downtime can be minimised using efficient and effective maintenance. To this end, repair-by-replacement of failed components and modules is useful. Failed components are replaced by spares and repaired off-line. Because many components of technically advanced systems are (very) expensive, it is often profitable to repair them instead of scrapping them. Therefore, the maintenance time is influenced by the number of spares and the repair shop capacity. We have a trade-off: the need for spares is reduced (reducing costs) using sufficient repair capacity (increasing costs) and the other way round. Also the preventive maintenance policy is relevant. Frequent preventive maintenance is costly (e.g. due to set-ups), reduces the risk of system failures and causes a more steady workload of failed components to the repair shop, yielding shorter throughput times (thereby reducing the need for spares and costs). The throughput times through

*Corresponding author. Tel.: +31 70 374 01 13; fax: +31 70 734 06 42.
E-mail address: karin.desmidt@tno.nl (K.S. de Smidt-Destombes).

0925-5273/$ - see front matter
doi:10.1016/j.ijpe.2006.08.013

a repair shop can also be reduced using a proper repair job scheduling mechanism (priority setting).

We see that the system availability is influenced in many ways, and that it is important to balance control variables like the amount of spares, maintenance frequency, repair shop capacity and repair shop scheduling. Until now, little research has been done about such integral trade-offs. In this paper we present a model to analyse these relations for an installed base of identical k-out-of-N systems sharing spares and repair capacity. A k-out-of-N system consists of N identical components of which at least k components are needed for a functional system. We focus on quantitative modelling and analyse the relation between control variables and system availability.

Fig. 1 shows two examples of systems that we consider in this paper. The one on the left is the Active Phased Array Radar (APAR), which consists of four sides, so called faces. Each face consists of several thousands of transmit and receive elements. Not all these elements have to be functioning in order to have the radar functioning. Therefore, this system can be considered as a k-out-of-N system. To give an indication, the APAR is approximately a 2700-out-of-3000 system. The Royal Netherlands Navy has multiple ships equipped with this radar. Therefore, there is an installed base sharing the same spares and repair capacity. This is also valid for the second system (Fig. 1 on the right), the Active Towed Array Sonar (ATAS). It consists of several tens of hydrophones used to detect objects beneath the water surface (like submarines). Not all hydrophones need to be functional to have the system functioning. To give an indication, we say this is a 58-out-of-64 system.

In this paper, we present approximation methods to set the maintenance interval, number of spares and repair capacity simultaneously for systems with and without component wear-out. In the case without component wear-out, a component is either working or failed (e.g. electronical components). We use exponential distributions for the component failure times, which are appealing because of their memoryless property. In case of component wear-out, we define component states: working, degraded and failed. Consider for instance the transmit and receive elements of the APAR. Components are fully operational if they can send and receive signals. If one of these functions fails, we call a component degraded, and failed if both functions fail. We assume that a component has an exponentially distributed time to move from "working" state to "degraded" state and an exponentially distributed time to move from "degraded" state to "failed" state.

Most literature about maintenance models only look at the maintenance policy, see e.g. the survey of Wang (2002). Usually, spares and repair capacity are assumed to be available and the relevant decisions concern only the maintenance policy such as intervals for inspections, maintenance (perfect, minimal, imperfect) and replacements, see e.g. (Abdel-Hameed, 1995). Sometimes, the action taken depend upon the number of failures as in the model

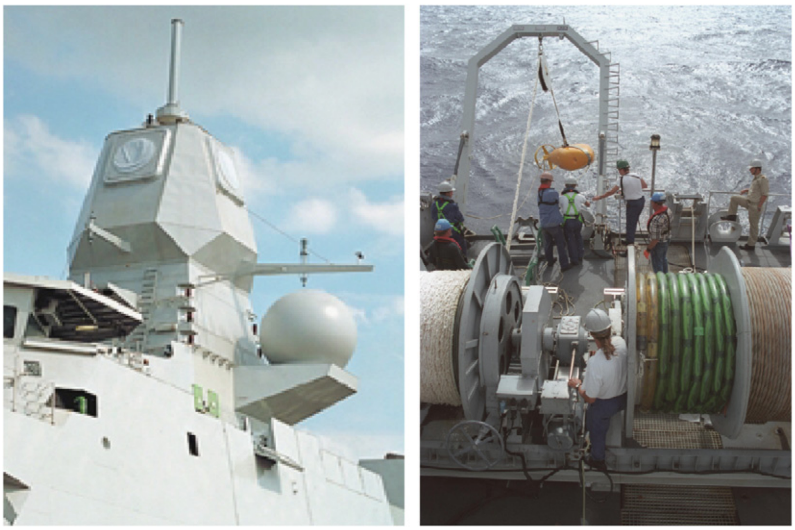

Fig. 1. *Left*: the Active Phased Array Radar system. *Right*: the Active Towed Array Sonar system.

presented by Love and Guo (1996) with Weibull failure rates.

There is only a limited amount of literature that mentions the importance of integrating the maintenance strategy with spares and repair capacity (e.g. Gross et al., 1985; Dinesh Kumar et al., 2000) but do not present a quantitative model. Bahrami-G et al. (2000) present a model to find the optimal length of the maintenance interval for equipment that deteriorates in time. However, dependency on the number of spares or repair capacity for the maintenance duration is not taken into account. Spare provisioning policy has been taken into account simultaneously with the maintenance policy by e.g. Kabir and Al-Olayan (1996), Kabir and Farrash (1996) and Park and Park (1986). In their models, maintenance strategy is an age based one and components are non-repairable. Chiang and Yuan (2001) for instance try to find an optimal inspection period combined with the best replenishment period and stock level. Brezavšček and Hudoklin (2003) present a model with a joint optimisation of a block replacement maintenance policy and spare parts policy. Again, the components are not repairable, which is encountered in most models that concern joint optimisation of a maintenance policy and a spares provisioning policy. The repair shop is obviously not modelled. In De Smidt-Destombes et al. (2004) a similar model is presented for a single system with no component wear-out under a condition based maintenance strategy. A closely related model is described in De Smidt-Destombes et al. (2006) considering one system with component wear-out. Compared to the last two models, this paper considers a block replacement policy instead of a condition based maintenance policy. The second difference is that the two models mentioned have spares and repair capacity for only a single system. In this paper, we consider several systems sharing spares and repair capacity. For a recent overview of repairable spare parts inventory analysis under finite repair capacity, we refer to Sleptchenko et al. (2002).

The outline of this paper is as follows. We start with the notation and the description of the basic model with no component wear-out (i.e. exponential time to failure) and exponential repair time of components in Section 2. In Sections 3 and 4, we, respectively, address the analysis of the model without component wear-out and with component wear-out. Section 5 shows the results of these approximation models compared with results of discrete event simulation. We end this paper with some conclusions and possibilities for further research in Section 6.

2. Notation and model description

2.1. Notation

Throughout this paper we use the following input parameters:

k	minimum number of system components needed
N	total number of system components
c	repair capacity
S	total number of spares
T	interval between two succeeding maintenance arrivals
M	size of the installed base
λ	component failure rate (without wear-out)
λ_1	transition rate from state 0 to 1 (with wear-out)
λ_2	transition rate from state 1 to 2 (with wear-out)
μ	component repair rate (without wear-out)
μ_1	repair rate from state 1 to 0 (with wear-out)
μ_2	repair rate from state 2 to 0 (with wear-out)

The next random variables are used:

$U(t)$	operational time of system with maintenance interval $t(=\min\{\widetilde{U},t\})$
$\widetilde{U}$	system time to failure
D	maintenance time
D^*	time to system failure
β	probability that maintenance time is larger than 0
A_i	number of components in state $i\ (=0,1,2)$ at system arrival
B_i	number of spares in state $i\ (=0,1,2)$ at system arrival

Notation for random variables used for the model without component wear-out:

$R(n)$	time needed to repair n components
$Z(t)$	number of repairs during time t

Notation for random variables used for the model with component wear-out:

B	total number of spares that need repair ($= B_1 + B_2$)
$T(i,j)$	time for the system transition from state $(N, 0, 0)$ to $(N-i-j, i, j)$
$R_i(n)$	time needed to repair n components from state i $(= 0, 1, 2)$
W_i	number of repairs during maintenance of type i $(= 0, 1, 2)$
$Z_i(t)$	number of type i repairs during time t

2.2. Model description

We consider an installed base of M identical k-out-of-N systems with hot stand-by redundancy. Hot stand-by redundancy means that all components, even the ones that are strictly not necessary, are operational and are therefore subject to failure with the same failure rate. Each component fails according to an exponential distribution with a failure rate λ per time unit. We assume that each system is maintained with a fixed maintenance interval of length T. In other words, we use a block replacement policy with no action taken if the system fails before its maintenance period. We assume that when a system has failed and less than k components are working, the system is not shut down. As an example, consider the APAR that can still work if less than 2700 out of the 3000 transmit-and-receive elements are available, although the performance is inferior (but better than nothing). Therefore, the components are still subject to failure after system failure. During maintenance, all failed components are replaced by spare components. The total number of spares for the installed base equals S. The components are repairable and are processed by a single repair shop with c identical, parallel repair channels. If the number of functional spares is insufficient to replace all failed components, the maintenance period is extended with the time needed to restore the lacking number of components. Repair of a failed component is exponentially distributed with a repair rate μ per time unit. The maintenance time, D, only consists of the waiting time for spares. We neglect the replacement time of components.

In Fig. 2, we show the various cycles that we distinguish when modelling the system. We have a *cycle for each system* in the installed base, defined as the period between two consecutive arrivals of the same system at the repair shop for maintenance (a fixed period with length T), and a *repair shop* cycle, defined as the period between the arrivals of two consecutive systems (a fixed period with length $\frac{T}{M}$). Both cycles start just before a system arrives for maintenance. The figure shows an example with an installed base $M = 3$ systems. The availability of each system is defined as the uptime of the system divided by the uptime plus the downtime, which equals $\frac{T-D}{T}$ if no system failure occurs during the operational time. Taking into account system failures and defining $U(T-D)$ as the uptime during $T-D$, the availability equals:

$$Av = \frac{E[U(T-D)]}{T}. \tag{1}$$

The maintenance duration D depends on the number of failed system components and the number of spares available at the start of maintenance as well as the repair capacity. Assuming that the failure rate and repair rate are known, we can control the system availability by the cycle length T, the number of spares S and repair capacity c.

For the analysis of this system, queueing models seem to be suitable at first sight. The repair shop can be modelled as a multi-server queue with batch arrivals, similar to the $D^X/M/c$ queue. The time between the arrivals of batches is deterministic (equal to $\frac{T}{M}$) and the number of components in each batch is a random variable that, unfortunately, depends on the system uptime $T-D$ and is therefore dependent on the repair shop performance. If the repair shop is highly utilised, the maintenance duration D increases, so the system uptime $T-D$ decreases and so the work offered to the system decreases. Theoretically, it is even possible that $D > T$, and then there are no failed components offered to the repair shop $M-1$ repair cycles later. As a consequence, the system is always stable having a utilisation of at most 1. Of course, the system availability is very low if the repair shop capacity is low. We also observe that it is not straightforward to estimate the repair shop utilisation in advance because of the relation between repair shop capacity and component arrival rate. Therefore, we have to use our approximations or simulations to estimate the repair shop utilisation. We conclude that the repair shop can be modelled as a non-standard queueing system for which no

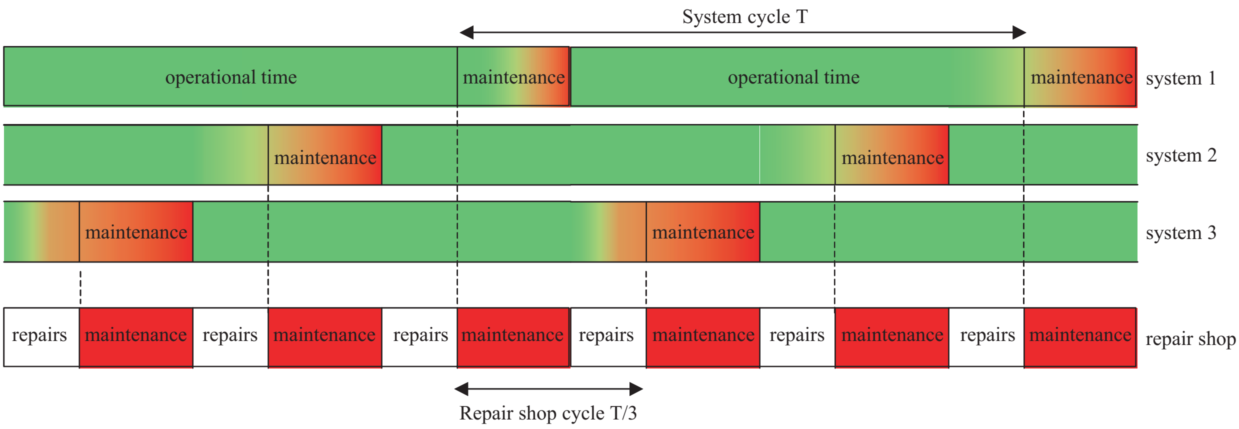

Fig. 2. A schematic representation for an installed base consisting of three systems.

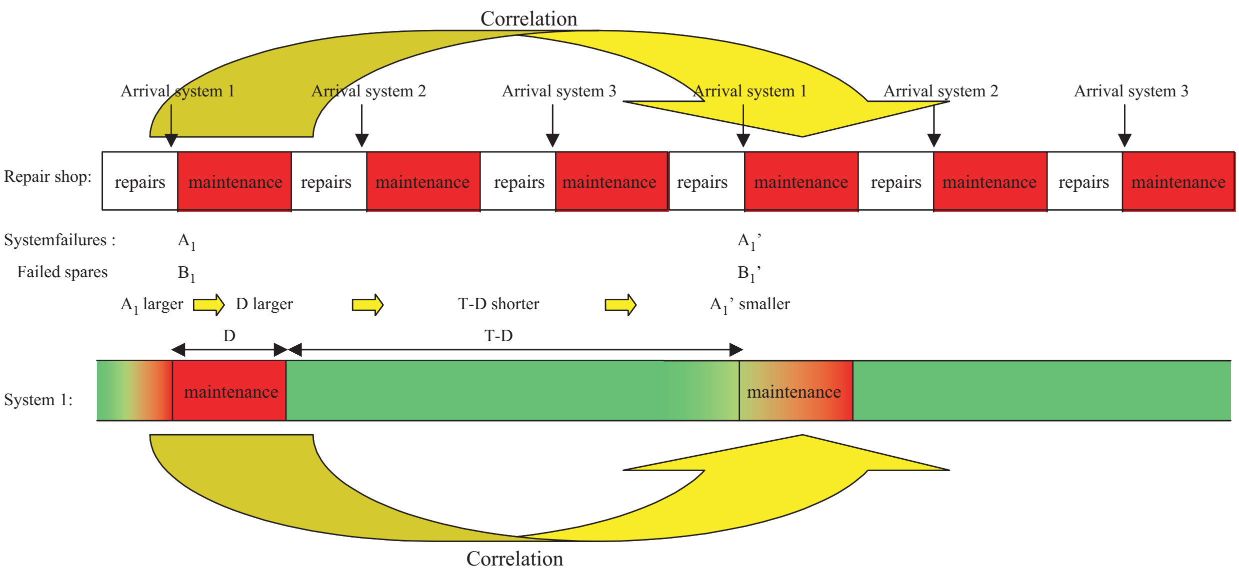

Fig. 3. Arrivals at the repair shop for an installed base of three systems.

suitable results are available in the literature to the best of our knowledge.

As another option, it seems to be logical to use renewal theory. However, we face the complication that consecutive system cycles are (possibly heavily) correlated, which induces correlations between repair shop cycles as well. We can explain this using Fig. 3 as follows. A k-out-of-N system arrives every T time units for maintenance. Maintenance is finished as soon as sufficient ready-for-use components are available to replace all failed components, which takes some time D (where $D=0$ if the number of functional spares is sufficient to replace all failed components immediately). The operational time in the next system cycle equals the time until the start of the next system maintenance, $T-D$. Now suppose that the system has more failed components than average, upon arrival for maintenance at the repair shop in the first system cycle. Then the maintenance duration D will probably be longer than average and so the operational time in the next cycle $T-D$ will be shorter than average. As a consequence, the number of failed components will be less than average when the system arrives again at the repair shop for maintenance in the second system cycle. Hence, we expect a negative correlation between the number of failed system components at the start of two consecutive system cycles for the same system in the installed base. From Fig. 3, we see that this also means a negative correlation between repair shop cycles, because the start of a cycle for each system in the installed base coincides with the start of a repair shop cycle. So, we expect a negative correlation between the number of failed components arriving at the repair shop in one cycle and the number of failed system components arriving M repair shop cycles later. This correlation is very hard to quantify. Therefore, we ignore this correlation in our model and assume that both the repair shop cycles and the system cycles are mutually independent. In the numerical section, we see to which extent this assumption has a significant impact on the accuracy of our approximations by comparison to results from discrete event simulation.

In the next sections, we focus on the approximation of $E[U(T-D)]$ given the independence assumption as stated above. In Section 3, we address the simple case without component wear-out (the component time to failure is exponentially distributed). We derive a set of stochastic equations for the maintenance duration D. We present two approximation methods to solve the system of equations for D based on the first two moments of the key random variables involved. The first approximation is based on continuous probability distributions (particularly suitable for large systems) and the second approximation is based on discrete probability distributions (particularly suitable for small systems).

In Section 4, we extend our model to include wear-out.

3. Analysis without component wear-out

As stated in Section 2.2 we need to determine the expected uptime $U(T-D)$ of the system during the operational time $T-D$. In the remainder of this paper, we simply use the shorthand notation U. If the system is still operational when it arrives for maintenance (i.e. the number of failed components is at most $N-k$), we have that $U=T-D$. However, if the system fails before maintenance starts, the uptime equals the time until system failure. Let us use $\widetilde{U}$ to denote the system time to failure if there is no maintenance. Then we can write $U=\min\{T-D,\widetilde{U}\}$. It is easy to find $\widetilde{U}$ as we show at the end of Section 3.1. The unknown variable we focus on first is the maintenance duration D. Before we do so, we give a list of the assumptions we use throughout this section.

1. All components have the same exponentially distributed time to failure.
2. The failure behaviour of the components is independent of each other.
3. There are no component failures during maintenance activities, as the system is down.
4. During maintenance all servers c are continuously busy (which is always true when the number of servers is less than the number of spares).
5. The same assumption is made for the time between two system arrivals.
6. Consecutive system cycles are independent.
7. Consecutive repair shop cycles are independent.

Now let us derive stochastic equations for the maintenance duration D based on the repair shop cycle. We define A_1 as the number of failed components in the system that arrives for maintenance at the start of the repair shop cycle. Also, we define B_1 as the number of failed components waiting for repair at the start of the same repair shop cycle, see Fig. 3. If there is no other system in repair, we have that $B_1 \leqslant S$. If at least one other system is still in repair, $B_1 > S$ (all spares are failed and there are some additional failed components from systems that arrived in the preceding repair shop cycles that have not been repaired yet). If $A_1+B_1 \leqslant S$, the number of ready-for-use spares is sufficient to replace all failed components immediately and hence the repair time is zero. If $A_1+B_1 > S$, the maintenance duration equals the time needed to restore A_1+B_1-S failed components. Denoting the time to restore X components as $R(X)$ and using the notation $X^+=\max\{X,0\}$ for any variable X, we write for D

$$D = R([B_1 + A_1 - S]^+). \tag{2}$$

We find a stochastic equation for B_1 (using assumption 4) by noting that the number of failed components at the start of a repair cycle equals the number of failed spares from the previous cycle plus the number of failed components from the system that arrived the previous cycle minus the number of spares restored between the two system arrivals (repair cycle with length $\frac{T}{M}$). In a stable situation, the probability distribution of B_1 should be identical at the start of all repair cycles. So if we define $Z(X)$ as the number of spares repaired during a period with length X, we find the stochastic equation

$$B_1 = \left[B_1 + A_1 - Z\left(\frac{T}{M}\right)\right]^+. \quad (3)$$

Next we have A_1, which is the number of failed system components during $T - D$

$$A_1 = \text{number of failed components during } T - D. \quad (4)$$

Conditioning on $T - D$, A_1 has a binomial distribution with parameters N and $1 - e^{-\lambda(T-D)}$, because the probability of a component failure during $T - D$ equals $1 - e^{-\lambda(T-D)}$.

In theory, we can find the probability distributions of A_1, B_1 and D by solving the set of stochastic equations (2)–(4). Unfortunately, an analytical solution is in general hard to find because of the complexity of these equations. A solution to our problem can be found in using the moment iteration approach as has initially been suggested by De Kok (1989) to approximate the waiting time in the $G/G/1$ queue from Lindley's equation. The moment iteration model we use to solve the set of equations is given in Section 3.1. For specific details of the moment iteration method, we refer to Section 3.2.

3.1. The moment iteration scheme

The moment iteration method is suitable to solve an implicit stochastic equation of the form $X = f(X)$, where $f(.)$ is some arbitrary function and X is some stochastic variable. The idea is to approximate the distribution of X by fitting a convenient distribution to the first two moments of the random variable X. In each iteration, we calculate improved estimates for the first two moments of X from the equation $X = f(X)$ using a two-moment approximation. We continue until the estimates for the first two moments of X do not change significantly anymore. We can do this, if it is relatively easy to calculate the first two moments of $f(X)$ for some specific family of probability distributions (e.g. Normal or Erlang distributions). This is particularly true for simple but common functions like $f(X) = \max\{X - C, 0\}$ and $f(X) = \max\{C - X, 0\}$ for some constant C. Although convergence cannot be proven, the moment iteration approach appears to converge in many practical situations, see for example Van der Heijden et al. (2001).

We can apply the same principle to a set of stochastic equations as we have here. We start with some arbitrary initial values for the first two moments of several random variables, approximate their distributions using a two-moment fit and generate improved approximations for the first two moments of the random variables involved, repeating this procedure until convergence. Again, we can do this if it is relatively easy to calculate the first two moments of some function $f(X, Y)$ for some specific family of probability distributions (e.g. Normal or Erlang distributions), particularly for simple but common functions like $f(X, Y) = \max\{X - Y, 0\}$.

Our iteration scheme to find the expected maintenance duration D involves two other key stochastic variables, A_1 and B_1, for which we use the set of equations given in (2)–(4). To find the mean and variance of $R(X)$ we use assumption 2. The conditional probability distribution of $R(X)$ (given X) has an Erlang distribution with X phases and scale parameter $c\mu$. Using the formulas for the conditional mean and variance, we find Eqs. (5) and (6). These expressions are used as an approximation, since it will not always be true that all servers are busy during the whole time $R(X)$. However, as long as there is not a surplus of capacity the c servers will be busy most of the time and this assumption is reasonable

$$E[R(X)] \approx \frac{E[X]}{c\mu}, \quad (5)$$

$$var[R(X)] \approx \frac{E[X]}{(c\mu)^2} + \frac{var[X]}{(c\mu)^2}. \quad (6)$$

In Eq. (3) for B_1, we defined $Z(X)$ for which we also use assumption 3, which gives us a Poisson distribution with parameter $c\mu X$. Hence, we find that the mean and variance are approximately given by

$$E\left[Z\left(\frac{T}{M}\right)\right] \approx c\mu\frac{T}{M}, \quad (7)$$

$$var\left[Z\left(\frac{T}{M}\right)\right] \approx c\mu\frac{T}{M}. \quad (8)$$

Finally, we need an expression for A_1 or $A_0 = N - A_1$. Because of assumption 1, the conditional distribution of A_0 given the length of the previous maintenance duration D is a binomial distribution with parameters N and $e^{-\lambda(T-D)}$. Similarly, the conditional distribution of A_1 given D is a binomial distribution with parameters N and $(1 - e^{-\lambda(T-D)})$. Hence,

$$E[A_0] = Ne^{-\lambda T}E[e^{\lambda D}], \tag{9}$$

$$\begin{aligned} var[A_0] &= E[var[A_0|D]] + var[E[A_0|D]] \\ &= E[A_0] + N(N-1)e^{-2\lambda T}E[e^{2\lambda D}] \\ &\quad - (E[A_0])^2. \end{aligned} \tag{10}$$

Directly determining $E[e^{\lambda D}]$ by fitting a continuous distribution for D is not very precise, due to the point mass in $D = 0$. Therefore, we define D^* as the maintenance duration, given that the maintenance duration is larger than zero. Hence, with $\beta = \Pr(A_1 + B_1 > S)$ and $E[e^{\lambda D^*}]$ the Laplace transform of D^*,

$$E[e^{\lambda D}] = (1-\beta) + \beta E[e^{\lambda D^*}]. \tag{11}$$

Our moment iteration scheme to find the mean maintenance duration consists of the following steps:

Step 0. Initialisation, choose starting values for $E[A_1]$, $var[A_1]$, $E[B_1]$, $var[B_1]$, $E[D]$ and $var[D]$.

Step 1. Determine the first and second moment of A_1 using $E[A_1] = N - E[A_0]$ and $var[A_1] = var[A_0]$ and Eqs. (9) and (10) with (11).

Step 2. Fit a distribution to $X = A_1 + B_1$ using the new values of $E[A_1]$ and $var[A_1]$ that we found in step 1, assuming that A_1 and B_1 are independent.

Step 3. Determine the first and second moment of $B_1 = [X - Z(\frac{T}{M})]^+$ with the mean and variance of $Z(\frac{T}{M})$ as given in Eqs. (7) and (8).

Step 4. Find the first and second moment of $[X - S]^+$ with $X = A_1 + B_1$ using the new values of $E[B_1]$ and $var[B_1]$ that were found in the previous step.

Step 5. Approximate the first and second moment of $D = R([X - S]^+)$ using Eqs. (5) and (6).

Step 6. Convergence check. If the relative difference between the $E[D]$ found in this iteration and the previous one is smaller than some fixed ε then stop, else go to step 1. In our model we chose $\varepsilon = 10^{-5}$.

The impact of the initial values on $E[D]$ is discussed in Section 5. After finding an approximation for the maintenance duration, we still need to find the mean operational time $E[U]$. We define the operational time as

$$U = \min\{T - D, \widetilde{U}\} = T - D - [T - D - \widetilde{U}]^+. \tag{12}$$

As stated earlier we can determine relatively easily the first two moments of $[X - Y]^+$ with X and Y positive random variables. Therefore, we define a positive random variable $X = T - D$ and fit a distribution to X and to $\widetilde{U}$ and find the mean of $[T - D - \widetilde{U}]^+$. $E[U]$ then equals $E[T - D] - E[[T - D - \widetilde{U}]^+]$.

We therefore need the mean and variance of D, which we determine using our iteration scheme, and we need the mean and variance of $\widetilde{U}$. $\widetilde{U}$ is the sum of the interval until the first component failure and the interval between the first and second failure,..., until the interval between failures $N - k$ and $N - k + 1$. The mean $\widetilde{U}$ equals the sum of the mean interval lengths and the variance of $\widetilde{U}$ equals the sum of the variances of the interval lengths

$$E[\widetilde{U}] = \sum_{i=k}^{N} \frac{1}{i\lambda}, \tag{13}$$

$$var[\widetilde{U}] = \sum_{i=k}^{N} \frac{1}{(i\lambda)^2}. \tag{14}$$

3.2. Large systems versus smaller systems

In the moment iteration scheme as presented in Section 3.1, we need the Laplace transform of D^* and we need to fit distributions. Therefore, we distinguish systems with a small number of components and systems with a large number of components. For large systems (systems like the APAR) we are able to use an Erlang distribution (see Tijms, 1994), while for smaller ones (systems like the ATAS) we use some specific discrete distributions. Dependent on the first two moments, we either use a mixture of two binomial distributions, a mixture of two negative binomial distributions, a mixture of two geometric distributions or a Poisson distribution, see Adan et al. (1995).

For systems with a large number of components, $E[e^{\lambda D^*}]$ is the Laplace transform of D^* for which we use an Erlang distribution with parameters

$\alpha = \frac{E[D^*]}{var[D^*]}$ and $r = \frac{(E[D^*])^2}{var[D^*]}$. This results in the following expression for the Laplace transform of D^*:

$$E[e^{\lambda D^*}] = \int_{t=0}^{T} e^{\lambda t} f_{D^*}(t)\,dt$$
$$= \begin{cases} (-1)^r\left(\frac{\alpha}{\lambda-\alpha}\right)^r \\ \quad +\alpha^r \sum_{i=0}^{r-1} \frac{-1^{i+r-1} T^i e^{(\lambda-\alpha)T}}{i!(\lambda-\alpha)^{r-i}}, & \alpha<\lambda, \\ \frac{(\alpha T)^r}{r!}, & \alpha=\lambda, \\ \left(\frac{\alpha}{\alpha-\lambda}\right)^r - \alpha^r \sum_{i=0}^{r-1} \frac{T^i e^{-(\alpha-\lambda)T}}{i!(\alpha-\lambda)^{r-i}}, & \alpha>\lambda. \end{cases} \tag{15}$$

For smaller systems, we use one of the discrete distributions as mentioned above.

- If the distribution of D^* is approximated by a mixture of two binomial distributions: $q\text{Bin}(n,p)+(1-q)\text{Bin}(n+1,p)$ the mean and variance of A_0 become

$$E[A_0] = Ne^{-\lambda T}\Big((1-\beta)+\beta\big(q\big((1-p)pe^{\lambda}\big)^n +(1-q)((1-p)pe^{\lambda})^{n+1}\big)\Big),$$

$$var[A_0] = E[A_0]-(E[A_0])^2 + N(N-1)e^{-2\lambda T}((1-\beta) + \beta\big(q((1-p)pe^{2\lambda})^n +(1-q)((1-p)pe^{\lambda})^{n+1}\big)).$$

- If the distribution of D^* is approximated by a mixture of two negative binomial distributions: $q\text{NegBin}(n,p)+(1-q)\,\text{NegBin}(n+1,p)$ the mean and variance of A_0 become

$$E[A_0] = Ne^{-\lambda T}\left((1-\beta)+\beta\left(q\left(\frac{p}{1-(1-p)e^{\lambda}}\right)^n +(1-q)\left(\frac{p}{1-(1-p)e^{\lambda}}\right)^{n+1}\right)\right),$$

$$var[A_0] = E[A_0]-(E[A_0])^2 + N(N-1)e^{-2\lambda T} \times\left((1-\beta)+\beta\left(q\left(\frac{p}{1-(1-p)e^{2\lambda}}\right)^n +(1-q)\left(\frac{p}{1-(1-p)e^{2\lambda}}\right)^{n+1}\right)\right).$$

- If the distribution of D^* is approximated by a mixture of two geometric distributions: $q\text{Geo}(p_1)+(1-q)\,\text{Geo}(p_2)$ the mean and variance of A_0 become

$$E[A_0] = Ne^{-\lambda T}\left((1-\beta)+\beta\left(q\frac{p_1}{1-(1-p_1)e^{\lambda}} +(1-q)\frac{p_2}{1-(1-p_2)e^{\lambda}}\right)\right),$$

$$var[A_0] = E[A_0]-(E[A_0])^2 + N(N-1)e^{-2\lambda T} \times\left((1-\beta)+\beta\left(q\frac{p_1}{1-(1-p_1)e^{2\lambda}} +(1-q)\frac{p_2}{1-(1-p_2)e^{2\lambda}}\right)\right).$$

- If the distribution of D^* is approximated by a Poisson distribution: $\text{Pois}(v)$, equations of the mean and variance of A_0 become

$$E[A_0] = Ne^{-\lambda T}((1-\beta)+\beta e^{v(e^{\lambda}-1)}),$$
$$var[A_0] = E[A_0]-(E[A_0])^2 + N(N-1)e^{-2\lambda T}((1-\beta)+\beta e^{v(e^{2\lambda}-1)}).$$

4. Model with ageing of components

We model a wear-out process using three component states 0–2 for a fully functional, degraded and failed component, respectively. Again we use the assumptions given in Section 3 and:

1. State transitions from state 0 to state 1 occur according to an exponential distribution with rate λ_1.
2. State transitions from state 1 to state 2 occur according to an exponential distribution with rate λ_2.
3. There are no direct transitions possible from state 0 to state 2.
4. During maintenance all degraded and failed components are replaced by spare components.
5. The repair times for degraded and failed components are exponentially distributed with rates μ_1 and μ_2, respectively.

To derive approximations for this model, we use an intermediate step, namely the special case where the repair rates of degraded and failed components are identical, $\mu_1 = \mu_2$ (Section 4.1). Next, we address the variant where the repair rates may be different (Section 4.2). In the latter case, it makes a difference in which order we repair degraded and failed components because of the different repair rates, $\mu_1 \neq \mu_2$. Hence, we can use a scheduling rule to decide in which order the spares are restored.

4.1. Degraded and failed components with equal repair rates

If the repair rates of the degraded and failed components are equal, it is sufficient to know the total number of components that are waiting or in repair at the start of a repair cycle, which we define as B. The total number of failed and degraded components at the start of a repair cycle equals $B + A_1 + A_2$ with A_i is the number of system components in state i at system arrival in the repair shop ($i = 0, 1, 2$). Then our equation for the maintenance duration (2) changes to

$$D = R([B + A_1 + A_2 - S]^+) = R([B + N - A_0 - S]^+). \tag{16}$$

Similar to the model without wear-out, the conditional distribution of the number of components in state 0 at the start of a repair cycle given the length of the previous maintenance time, $A_0|D$, is binomial with parameters N and $e^{-\lambda_1(T-D)}$. The unconditional mean and variance of this stochastic variable are given by Eqs. (9) and (10) with λ replaced by λ_1.

The number of components that are not yet restored at the start of the repair cycle B equals the number of spares to restore at the previous system arrival, plus the failed components that came out of that system minus the number of repairs that is done between the two system arrivals. Hence, we have the same equation as for the model without wear-out, see Eq. (3) with $A_1 = N - A_0$

$$B = \left[B + N - A_0 - Z\left(\frac{T}{M}\right)\right]^+. \tag{17}$$

Analogous to the model without ageing of components we use a moment iteration method to solve this set of Eqs. (16), (17) and (9)–(11). The generic iteration scheme is as follows:

Step 0. Initialisation, choose start values for $E[A_0]$, $var[A_0]$, $E[B]$, $var[B]$, $E[D]$ and $var[D]$. Approximate the mean and variance of $Z(\frac{T}{M})$ as in Eqs. (7) and (8).

Step 1. Determine the mean and variance of A_0 using Eqs. (9)–(11).

Step 2. Find the mean and variance of $Y = N - A_0 + B$, assuming that A_0 and B_1 are independent.

Step 3. Find the mean and variance of $B = [Y - Z(\frac{T}{M})]^+$.

Step 4. Find the mean and variance of $X = [Y - S]^+$ after determining the mean and variance of $Y = N - A_0 + B$ again with the new values for the mean and variance for B found in step 3.

Step 5. Approximate the mean and variance of the maintenance duration $D = R(X)$ using Eqs. (5) and (6).

Step 6. Convergence check. If the relative difference between the $E[D]$ found in this iteration and the previous one is smaller than some fixed ε then stop, else go to step 1. In our model, we chose $\varepsilon = 10^{-5}$.

To find the mean operational time $E[U]$, we also need the mean and variance of $\widetilde{U}$, which changes because of the ageing of components. In De Smidt-Destombes et al. (2006) a recursive method is presented to find the first two moments. In short, this method works as follows. We define $T(i,j)$ as the time duration to get from state (i,j) to a failed state which has $N - k + 1$ components in state 2. State (i,j) refers to $N - i - j$ components in state 0, i components in state 1 and j components in state 2. This immediately gives us the starting values of the recursion, $T(i, N - k + 1) = 0$ and $T^2(i, N - k + 1) = 0$ for every value of $0 \leqslant i \leqslant k - 1$. The recursion is given by

$$E[T(i,j)] = \tau(i,j) + \alpha(i,j)E[T(i+1,j)] + \beta(i,j)E[T(i-1,j+1)],$$

$$E[T^2(i,j)] = 2\tau(i,j)E[T(i,j)] + \alpha(i,j)E[T^2(i+1,j)] + \beta(i,j)E[T^2(i-1,j+1)].$$

Here, $\tau(i,j) = \frac{1}{(N-i-j)\lambda_1 + i\lambda_2}$ is defined as the expected sojourn time in state (i,j), $\alpha(i,j) = \frac{(N-i-j)\lambda_1}{(N-i-j)\lambda_1 + i\lambda_2}$ is the probability of a transition from state 0 to state 1 and $\beta(i,j) = \frac{i\lambda_2}{(N-i-j)\lambda_1 + i\lambda_2}$ is the probability of a transition from state 1 to state 2. Now $\widetilde{U}$ is defined as the time from state $(0,0)$ to a failed state. Hence,

$$E[\widetilde{U}] = E[T(0,0)],$$

$$var[\widetilde{U}] = E[T^2(0,0)] - (E[T(0,0)])^2.$$

With the first two moments of $\widetilde{U}$ and D, we are able to determine the expected uptime U from Eq. (12).

4.2. Degraded and failed components with different repair rates

4.2.1. Repair strategy

We denote the repair rate for degraded and failed components by μ_1 and μ_2, respectively. It is plausible that repair of failed components takes more time on average than repair of degraded components, so $\mu_1 \geqslant \mu_2$. The remainder of our analysis is based on this assumption for ease of notation. It is straightforward to modify the analysis if $\mu_2 > \mu_1$.

If the two repair types (degraded and failed components) have different repair rates, we can influence the maintenance duration and hence the availability by choosing the order in which the repair jobs are processed. Hence, we have an additional degree of freedom, namely a repair priority rule that we can use to minimise the maintenance duration. We know that we should recover exactly $[B+A_1+A_2-S]^+$ components to restore the system that arrived at the start of a repair cycle. Therefore, we have to choose (a) how many of these $[B+A_1+A_2-S]^+$ components should be degraded components and how many should be failed components (b) in which order are we going to repair these $[B+A_1+A_2-S]^+$ components. Regarding the first issue, it is obvious that we should select as many degraded components as possible, because their repair rate is higher. If we have insufficient degraded components, we add failed components until we reach the required number of $[B+A_1+A_2-S]^+$ components. Regarding the second issue, we can use the fact that we can minimise the makespan of a fixed set of repair jobs by selecting the longest processing times first, see for instance Pinedo and Chao (1999). So, within the set of degraded and failed components that we should repair to recover the system, we should give priority to failed components. Summarised, our repair strategy is as follows:

> If the number of degraded spares (state 1) is sufficient to replace all components in the system, then only repair degraded ones. If the number of degraded spares is not sufficient, start repairing the minimum number of failed components needed to repair the system and next repair all degraded components.

During the time in which the repair shop repairs components without a direct demand (the periods between maintenance periods in Fig. 2) we want the repair shop to restore as many spare parts as possible. Therefore, during this time the priority rule is:

> First repair all degraded spares and then start repairing failed spares.

Using these priority rules, we are able to define a set equations, which is presented in Section 4.2.2.

4.2.2. Model with wear-out and different repair rates

The approach to the problem with ageing of components and different repair rates is analogous to the one with distinguishing the components in states 1 and 2. The maintenance duration is therefore splitted into two parts, one part for the number of type 1 repairs and one part for the type 2 repairs. We define W_1 and W_2 as stochastic variables for the number of repairs of type 1 and repairs of type 2, respectively, during the maintenance time. Then, we approximate the expected maintenance duration and its variance by

$$E[D] \approx \frac{E[W_1]}{c\mu_1} + \frac{E[W_2]}{c\mu_2}, \tag{18}$$

$$var[D] \approx \frac{E[W_1] + \mathrm{var}[W_1]}{(c\mu_1)^2} + \frac{E[W_2] + \mathrm{var}[W_2]}{(c\mu_2)^2}. \tag{19}$$

To determine the workload of types 1 and 2 we use the priority rule as discussed in the previous section. This implies the workload of type 2 components to be zero as long as the total number of failed components, $A_2 + B_2$, is at most equal to S. Otherwise the workload is equal to the difference between the total number of failed components and the number of spares

$$W_2 = [A_2 + B_2 - S]^+. \tag{20}$$

For the workload of type 1 components we consider the total workload and subtract the workload of type 2 components. The total number of degraded and failed components in the system and the repair shop equals $A_1 + A_2 + B_1 + B_2 = N - A_0 + B_1 + B_2$. The total workload is the total number of degraded and failed components minus the number that does not need to be restored during the system maintenance period. In other words, if the total number of components to restore is less than or equal to S, the total workload during maintenance is zero, otherwise the workload is the difference between the components to restore and S. Hence, we find for the workload of types 1 and 2 components

$$W_1 = [N - A_0 + B_1 + B_2 - S]^+ - W_2. \tag{21}$$

Because of the different repair rates, we split the number of unrestored spares B into B_1 and B_2 where B_i denotes the number of components in state i in the repair shop at arrival of a system. The number of spare components in state 1 is equal to the total number of components in state 1 just after the previous system arrival $(B_1 + A_1)$ minus the number of repairs done in the time left after the type 2 repairs W_2. For B_2, we assume that W_2 is repaired before the next system arrives. This is a reasonable assumption since these are the first components to be restored and the availability requirements of the systems are rather high. Then B_2 equals the total number of components in state 2 minus W_2 minus the number of restores done in the time left after W_2 and $B_1 + A_1$ are repaired. We then find the following equations for B_i:

$$B_0 = [S - B_1 - B_2]^+, \tag{22}$$

$$B_1 = \left[B_1 + A_1 - Z_1\left(\left[\frac{T}{M} - R_2(W_2)\right]^+\right)\right]^+, \tag{23}$$

$$B_2 = \left[B_2 + A_2 - W_2 - Z_2 \times\left(\left[\frac{T}{M} - R_2(W_2) - R_1(A_1 + B_1)\right]^+\right)\right]^+. \tag{24}$$

Here $R_i(X)$ is defined as the time needed to restore X components of type i and $Z_i(Y)$ is defined as the number of repairs of type i during time Y. If we use a moment iteration scheme to determine the maintenance duration using Eqs. (18)–(24) the results are not very good. In the expression for W_1, we have a correlation between the total workload during maintenance and the workload of type 2 components that we ignore in our approximations. This affects the variance of W_1 and consequently it also affects the variance of D. The simulations that we describe more detail in Section 5 showed that this correlation is often close to 1. For the approximation of B_2 we have a correlation between W_2 and $A_2 + B_2$ which is at least 0.6 according to our simulation. To deal with these problems, we can try to estimate the magnitude of the correlations. Unfortunately, this is mathematically hard. As an alternative, we can reformulate Eqs. (18)–(24) in terms of other random variables, such that the correlations are less severe. Below, we derive such alternative expressions for W_1 and B_2.

Regarding W_1 we know that if the maintenance duration equals zero, then the value of W_1 equals zero. The probability that the maintenance duration is larger than zero, is

$$\Pr(A_1 + A_2 + B_1 + B_2 > S) = \Pr(N - A_0 + B_1 + B_2 > S) = \beta.$$

The total number of spares to restore during the maintenance period is $N - A_0 + B_1 + B_2 - S$, under the condition that $N - A_0 + B_1 + B_2 - S > 0$. The time needed to restore this number of spares equals D^*. Now there are two possibilities. The first possibility is that we need to restore only part of the components in state 1. Then the value of W_1 becomes equal to the number of restores that can be done during D^*, $Z_1(D^*)$. The second possibility is that we need to restore all components in state 1 and maybe even a number of components in state 2. The value of W_1 is then equal to $A_1 + B_1$. Combining these different possibilities we find

$$W_1 = \min\{Z_1(D^*), A_1 + B_1\}\beta + 0(1 - \beta). \tag{25}$$

Regarding B_2, we add the assumption that we are also able to restore all type 1 components, $A_1 + B_1$, before the next system arrives in the repair shop. This assumption is not an unreasonable one as long as we are dealing with utilisation rates of the repair shop that are not too large, say 90–95%. Hence, we approximate B_2 by using the following expression:

$$B_2 = \left[B_2 + A_2 - Z_2\left(\left[\frac{T}{M} - R_1(A_1 + B_1)\right]^+\right)\right]^+. \tag{26}$$

For the mean and variance of A_0 we use the previous expressions (9) and (10). The number of components in state 1 also has a binomial distribution with parameters N and $\frac{\lambda_1}{\lambda_1 - \lambda_2}(e^{-\lambda_2(T-D)} - e^{-\lambda_1(T-D)})$ and the number of components in state 2 is binomially distributed with parameters N and $1 - e^{-\lambda_1(T-D)} - \frac{\lambda_1}{\lambda_1 - \lambda_2}(e^{-\lambda_2(T-D)} - e^{-\lambda_1(T-D)})$. With some algebra we find

$$E[A_1] = N\frac{\lambda_1}{\lambda_1 - \lambda_2}(e^{-\lambda_2 T}E[e^{\lambda_2 D}] - e^{-\lambda_1 T}E[e^{\lambda_1 D}]), \tag{27}$$

$$var[A_1] = (N^2 - N)\left(\frac{\lambda_1}{\lambda_1 - \lambda_2}\right)^2 (e^{-2\lambda_2 T} E[e^{2\lambda_2 D}] - 2e^{-(\lambda_1+\lambda_2)T} E[e^{(\lambda_1+\lambda_2)D}] + e^{-2\lambda_1 T} E[e^{2\lambda_1 D}]) + E[A_1] - (E[A_1])^2, \quad (28)$$

$$E[A_2] = N - E[A_0] - E[A_1], \quad (29)$$

$$var[A_2] = \frac{2N(N-1)\lambda_1}{\lambda_1 - \lambda_2}(e^{-(\lambda_1+\lambda_2)T} E[e^{(\lambda_1+\lambda_2)D}] - e^{-2\lambda_1 T} E[e^{2\lambda_1 D}]) + var[A_0] + var[A_1] - 2E[A_0]E[A_1]. \quad (30)$$

To find an approximation of the maintenance duration we use a moment iteration method using Eqs. (18)–(20), (22), (23), (25)–(30), (9)–(10). The iteration scheme becomes as follows:

Step 0. Initialisation, choose start values for $E[A_0]$, $var[A_0]$, $E[B_0]$, $var[B_0]$, $E[W_2]$, $var[W_2]$, $E[D]$ and $var[D]$.

Step 1. Determine the means and variances of A_0 (using Eqs. (9) and (10)) and A_1 and A_2 using Eqs. (27)–(30). Therefore, we again take out the point mass of D in zero and use Eq. (11) with $\beta = \Pr(A_1 + A_2 + B_1 + B_2 > S) = \Pr(N - A_0 + B_1 + B_2 > S)$.

Step 2. Find the mean and variance of $B_1 = [B_1 + A_1 - Z_1(X)]^+$ with $X = [\frac{T}{M} - R_2(W_2)]^+$. Therefore, we first determine the mean and variance of $R_2(W_2)$ using Eqs. (5) and (6) with μ replaced by μ_2 and $X = W_2$. Secondly, we find the mean and variance of the time available for type 2 repairs during a repair shop cycle: $X = [\frac{T}{M} - R_2(W_2)]^+$. Thirdly, we find the mean and variance of $Z_1(X)$ using the approximations given in Eqs. (7) and (8) with μ replaced by μ_1. Finally, we find the mean and variance of B_1.

Step 3. Find the mean and variance of $B_2 = [B_2 + A_2 - Z_2(Y)]^+$ with $Y = [\frac{T}{M} - R_1(A_1 + B_1)]^+$. Therefore, we first find the mean and variance of $R_1(A_1 + B_1)$ approximated by Eqs. (5) and (6) with $X = A_1 + B_1$ and μ replaced by μ_1. Secondly, we find the mean and variance of Y and thirdly, we find the mean and variance of $Z_2(Y)$ using the approximations given in Eqs. (7) and (8) with μ replaced by μ_2. Finally, we find the mean and variance of B_2.

Step 4. Find the mean and variance of $B_0 = [S - B_1 - B_2]^+$.

Step 5. Find the mean and variance of $W_1 = (A_1 + B_1 - [A_1 + B_1 - Z_1(D^*)]^+)\beta$ with β as found in step 1 and the mean and variance of $Z_1(D^*)$, with the mean and variance of D^* as we found in step 1 to take out the point mass.

Step 6. Find the mean and variance of $W_2 = [A_2 + B_2 - S]^+$.

Step 7. Approximate the mean and variance of the maintenance duration using Eq. (18) for $E[D]$ and Eq. (19) for $var[D]$.

Step 8. Convergence check. If the relative difference between the $E[D]$ found in this iteration and the previous one is smaller than some fixed ε then stop, else go to step 1. In our model we chose $\varepsilon = 10^{-5}$.

To compute the mean operational time for the systems $E[U]$ we use the same method as described in the model with ageing of components and equal repair rates.

In our model the maintenance duration is equal to zero as long as the number of components in the system that need to be replaced is smaller than or equal to the number of ready-for-use spares. This can be adjusted easily by adding the expected replacement time to the maintenance duration. Let us assume that v is the replacement rate of a component. Then the replacement time for a component is approximated by $\frac{1}{cv}$ and therefore the maintenance duration is increased by $\frac{EA_1+EA_2}{cv} = \frac{N-EA_0}{cv}$. Of course, one might argue that it is more reasonable to assume a deterministic replacement time, because component replacement is a well-defined task that usually shows little variation in the time required, unlike component repair. We refer to De Smidt-Destombes et al. (2004) for a model variant with deterministic replacement times.

5. Numerical results

5.1. Accuracy of the models

We constructed two models in this paper, one without component wear-out and a second one with component wear-out. For both types of systems convergence of the iteration scheme is found within roughly 10 iterations. Both models are approximations and we therefore need to check the accuracy of the models. To this end we constructed a discrete event simulation model in the object oriented simulation software eM-Plant 7.5 as a benchmark. In all cases, we simulated 1010 system cycles, where we ignored the first 10 cycles for output analysis (that is, we used a warm-up period of 10 system cycles). We used the batch means method (cf. Law and Kelton, 1991) to calculate a confidence interval for the mean availability and found that the half

Table 1
Combinations of input parameters used for systems with and without component wear-out

	M	$\lambda = \lambda_1$	λ_2	(T, S, c)	$\mu = \mu_1 = \mu_2$
7-out-of-10	2	0.0001	0.1	(2000, 4, 1), (2200, 4, 1), (2500, 4, 1), (3000, 4, 1)	0.0023–0.0036
	4	0.0001	0.1	(2000, 3, 1), (2000, 4, 2), (2500, 5, 1), (2750, 4, 1)	0.004–0.0075
	10	0.0001	0.05	(2000, 3, 1), (2750, 4, 1)	0.011–0.017
			0.1	(2000, 1, 1), (2000, 2, 1)	0.01125–0.0175
58-out-of-64	2	0.0001	0.1	(800, 4, 1) , (800, 6, 1), (900, 5, 1), (1000, 6, 1)	0.008–0.02
	4	0.0001	0.1	(800, 4, 1), (800, 6, 1), (900, 5, 1), (1000, 6, 1)	0.02–0.032
	10	0.0001	0.1	(800, 4, 3), (800, 6, 3), (900, 5, 3), (1000, 6, 3)	0.02–0.032
2700-out-of-3000	2	0.0001	0.1	(800, 250, 1), (800, 300, 1), (900, 275, 1), (1000, 300, 1)	0.4–1
	4	0.0001	0.1	(800, 250, 1), (800, 300, 1), (900, 275, 1), (1000, 300, 1)	1–1.6
	10	0.0001	0.1	(800, 250, 3), (800, 300, 3),(900, 275, 3), (1000, 300, 3)	1–1.6

Table 2
Mean (max) differences between the approximations and the simulation results for both the maintenance duration and the availability using discrete distributions and continuous distributions

	Without wear-out		With wear-out ($\mu_1 = \mu_2$)		With wear-out ($\mu_1 \neq \mu_2$)	
	$E[D]$	Av	$E[D]$	Av	$E[D]$	Av
Discrete						
7-out-of-10	22.2% (106.7%)	1.4% (11.7%)	19.4% (93.6%)	0.9% (7.7%)	14.0% (74.2%)	2.2% (7.6%)
58-out-of-64	3.5% (32.3%)	0.2% (2.1%)	5.1% (38.0%)	0.2% (2.5%)	4.1% (32.0%)	4.3% (7.1%)
2700-out-of-3000	–	–	–	–	–	–
Continuous						
7-out-of-10	29.4% (111.6%)	1.6% (13.6%)	63.2% (791%)	1.9% (19.0%)	32.7% (644%)	0.9% (12.5%)
58-out-of-64	4.0% (29.8%)	0.1% (0.6%)	5.0% (37.4%)	0.2% (2.5%)	4.6% (29.2%)	0.1% (1.0%)
2700-out-of-3000	3.5% (13.6%)	0.1% (0.5%)	3.7% (10.9%)	0.1% (0.4%)	2.5% (8.6%)	0.1% (0.4%)

width of the 95% confidence interval is (considerably) less than 1% in most cases.

We considered three different system sizes: 7-out-of-10, 58-out-of-64 and 2700-out-of-3000. For the latter one, which is a large system, we only used the approach with continuous distributions. For the other two system sizes we used the approximation with both discrete distributions and continuous distributions. The computation time for a 2700-out-of-3000 system is too large for the use of the approximation with discrete distributions.

In order to deal with realistic situations we consider systems with an availability of at least 90%. For a realistic utilisation rate of the repair shop we consider rates between 50% and 90%.

For each of the three system sizes we constructed about 80 combinations of values for the transition rates, repair rates, maintenance intervals, number of spares and number of repair capacity for both models, divided equally over the size of the installed base (see Table 1 for an overview of the parameter combinations used). For the model with component wear-out we chose $\mu_1 = \mu_2$. This gives us the opportunity to compare the model of Section 4.1 which requires equal repair rates and the more general model of Section 4.2 which does not require equal repair rates.

In Table 2 we show the mean and maximum relative differences that we found in the maintenance duration and in the availability compared to our simulation results. We did not find evidence that approximation errors depend on the repair shop utilisation or the system availability.

For large systems we have no choice other than to use an approximation with continuous distributions. Although, the approximation of the maintenance duration may not be very accurate at all times, we have a good approximation for the system availability. This is due to the fact that the availability of the systems is 90% or more and the maintenance duration is a relatively small part of the system's cycle. As a result the impact of an error in the maintenance duration is small.

For the medium sized systems without component wear-out we can choose between an approximation with continuous or discrete distributions. If we would split the results according to the size of the installed base we would see that for a larger installed base the approximation with discrete distributions is slightly better than the one with continuous distributions. For medium sized systems with component wear-out, and not necessarily with $\mu_1 = \mu_2$, we find that the results using continuous distributions are more accurate.

In case of the small systems we find a better performance if we use the approximation with discrete distributions. Including component wear-out and arranging the results according to the size of the installed base we see that for a large installed base it is best to use the approximation with continuous distributions and the one with discrete distributions for the smaller sizes of the installed base.

As seen in Table 2 the approximations for smaller systems are less satisfactory. This can be explained by the fact that an absolute small approximation error for A_i or B_i is a relatively large error when we only have a few components. As a result, the error in the approximation of the maintenance duration is relatively large. For systems with a larger number of components the relative approximation errors are therefore smaller. Without exception the maximum errors given in Table 2 are all generated by the scenarios with a smaller size of the installed base.

When the installed base is small the errors in the approximations are much worse than the errors we find for a larger installed base. This is probably due to the fact that there is a dependency between the cycles in which the same system arrives at the repair shop. If the installed base becomes larger this dependency becomes smaller because the number of intermediate cycles $(M-1)$ becomes larger. This is shown in Table 3, showing the results for the different number of systems per installed base for a 7-out-of-10 system.

In Fig. 4, we show the differences in approximation errors for maintenance duration as a function of the utilisation rate of the repair shop for the different sizes of the installed base consisting of

Table 3
Mean differences between the approximations and the simulation results for both the maintenance duration and the availability for a 7-out-of-10 system

# Systems per installed base	Without wear-out		With wear-out ($\mu_1 = \mu_2$)		With wear-out ($\mu_1 \neq \mu_2$)	
	$E[D]$ (%)	Av (%)	$E[D]$ (%)	Av (%)	$E[D]$ (%)	Av (%)
2 Systems	56.6	3.7	42.3	2.0	25.0	1.3
4 Systems	26.4	0.8	13.7	0.5	11.9	3.2
10 Systems	5.2	0.3	3.8	0.2	5.0	2.2

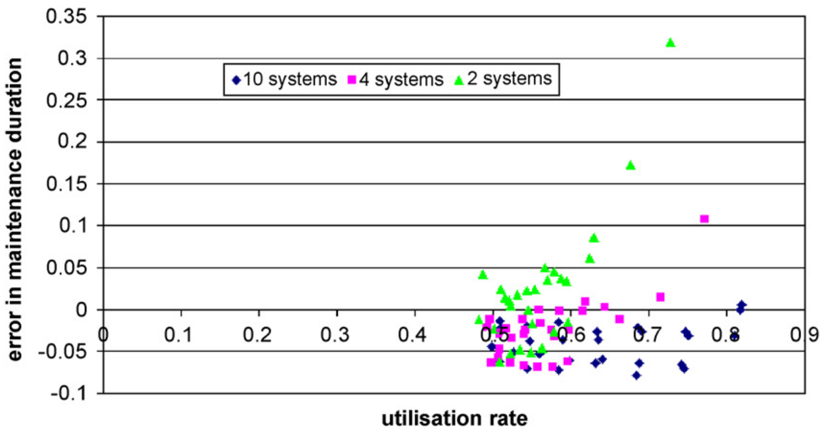

Fig. 4. Approximation errors in maintenance duration shown as a function of the utilisation rate of the repair shop.

58-out-of-64 systems. It is shown that the approximation errors become smaller as the size of the installed base increases. For larger systems, the errors for the maintenance duration are less and for smaller systems the differences tend to be larger.

Looking at Table 2 again, there is another interesting result. When we look at the approximation errors for the maintenance duration for the two models with component wear-out we see that the model that does not require the repair rates μ_1 and μ_2 to be equal to give less satisfactory approximations. While for the same scenarios we find that for the approximations of the availability the results are the other way around. The scenarios in which this happens are scenarios with either a small installed base or an availability level of over 99%. For the scenarios with a small installed base we already concluded that the model does not perform very well and for the scenarios with an availability level over 99% the absolute differences in the approximation errors are small.

So, for the remaining scenarios with smaller availability levels and a sufficiently large installed base the model that requires $\mu_1 = \mu_2$ outperforms the more general model with component wear-out.

5.2. Relations between decision variables T, S, c

In this section, we take a closer look at the relations between the decision variables T, S and c. Using an example we look at the effects that variations in the different variables have on the system availability and what trade-offs there are between these variables.

In Fig. 5 an example is shown for an installed base of ten 2700-out-of-3000 systems. If for instance the target availability level is 98%, we can see from the graph the different combinations of length of maintenance interval, number of spares and repair capacity, with which to achieve this availability level. Reducing the repair capacity can to a certain extent be compensated for by more frequent maintenance. For instance, with $c = 5$ and $S = 200$ we find an availability of 98.2% with a maintenance interval of 1050 time units. Bringing the capacity down to 4 or 3, we can achieve the same availability if we decrease the maintenance interval to 950 or 850 time units respectively. This confirms the expectations we mentioned in the introduction of this paper that a higher maintenance frequency leads to less variation in the component arrival process at the repair shop, so that less repair capacity is needed. For the number of spares we see similar results. Looking at it the other way around we see that with an increase of the spares from 200 to 300 we can increase our maintenance interval from 650 to 1100 and still have an availability of almost 99.5% with $c = 3$. So, with an increasing maintenance interval we can decrease the repair capacity, decrease the number of spares or decrease both. With this model the effects can be quantified for specific cases. Which combination of

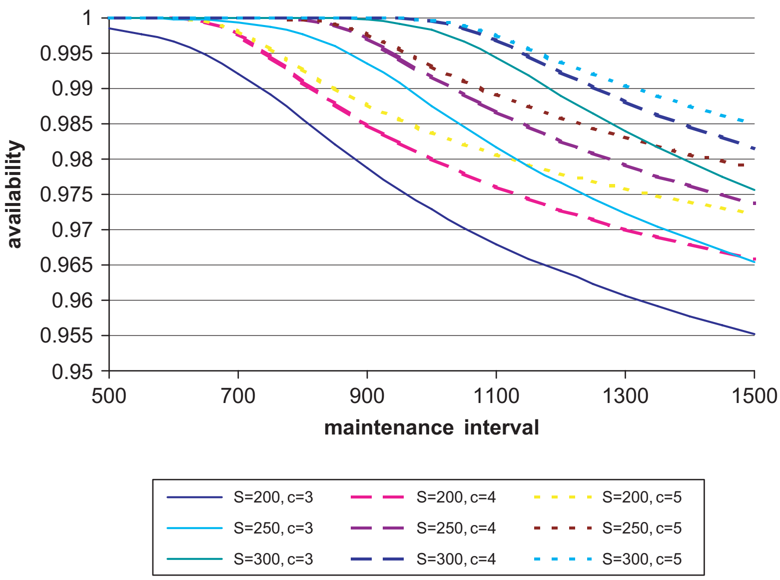

Fig. 5. Different combinations of maintenance interval length, number of spares and repair capacity can lead to similar availability levels.

parameters (T, S, c) is the best, depends on the cost involved. Without loss of performance the cheapest option can be chosen.

6. Conclusions and future research

From the previous section we conclude that we have accurate approximations for the availability as function of the maintenance interval, number of spare parts and the number of repair capacity, provided that the number of components in a system is not too small (say more than 10) and the size of the installed base is not too small (say at least 4). We can draw graphs in which we can quantify the effect of the length of the maintenance interval and the maintenance means (spares and capacity). Increasing the maintenance interval means we can do with less spares or capacity (or both) and still have the same availability performance. For systems with only a small number of components or a small installed base we have to be careful because the approximation errors may be relatively large.

In practice, multiple component types are usually restored in the same repair shop, sharing the same capacity. Then, we have to find a decision rule for the order in which items are repaired, especially, if the different items are all needed for the maintenance on a single system. Of each item, there has to be enough spares restored. The active phased array radar (APAR) system is again a good example. Besides the transmit and receive elements, the system consist of a large number of power supplies. To get the system fully operational during a maintenance period, we need transmit and receive elements and power supplies. The repair strategy during the maintenance period does not change very much. This duration is minimal as long as we take as many small repair jobs as possible, and perform the repair jobs from the longest to the shortest. The repair strategy for the period between the maintenance periods is less obvious. All sorts of items are needed, so a priority rule based on the repair length will not be optimal. What this rule should look like is subject to future research. Some research has been done on these priority rules in Sleptchenko et al. (2005).

If the assumption of system shutdown at $N-k+1$ failures is introduced we find smaller maintenance durations if the number of spares and capacity remains unchanged. The mean operational time $E[U]$ will increase if the value of T is not too large and as a result the system availability increases. To have the same availability as without system shutdown we may be able to lower the cost involved. Either by reducing the number of spares or capacity or by increasing the maintenance interval length.

Given the approximations as provided in this paper, we may want to find the optimal combination of maintenance interval and number of spares and repair capacity with respect to costs. Given the number of options for the decision variables and the computation times, enumeration is usually not an option. Therefore, the development of an optimisation method is a subject for further research.

References

Abdel-Hameed, M., 1995. Inspection, maintenance and replacement models. Computers and Operations Research 22 (2), 435–441.

Adan, I., Van Eenige, M., Resing, J., 1995. Fitting discrete distributions on the first two moments. Probability in the Engineering and Informational Sciences 9 (4), 623–632.

Bahrami-G, K., Price, J., Mathew, J., 2000. The constant-interval replacement model for preventive maintenance: a new perspective. International Journal of Quality & Reliability Management 17 (8), 822–838.

Brezavšček, A., Hudoklin, A., 2003. Joint optimization of block-replacement and periodic-review spare-provisioning policy. IEEE Transactions on Reliability 52 (1), 112–117.

Chiang, J., Yuan, J., 2001. Optimal maintenance policy for a markovian system under periodic inspection. Reliability Engineering and System Safety 71, 165–172.

De Kok, A., 1989. A moment-iteration method for approximating the waiting time characteristics of the G/G/1 queue. Probability in the Engineering and Informational Sciences 3, 273–287.

De Smidt-Destombes, K., Van Der Heijden, M., Van Harten, A., 2004. On the availability of a k-out-of-N system given limited spares and repair capacity under a condition based maintenance strategy. Reliability Engineering and System Safety 83 (3), 287–300.

De Smidt-Destombes, K., Van der Heijden, M., Van Harten, A., 2006. 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 (1), 182–200.

Dinesh Kumar, U., Crocker, J., Knezevic, J., El-Haram, M., 2000. Reliability, Maintenance and Logistic Support: A Life Cycle Approach. Kluwer Academic Publisher, Dordrecht, MA.

Gross, R., Miller, D., Soland, R., 1985. On common interests among reliability, inventory and queuing. IEEE Transactions on Reliability 34 (3), 204–208.

Kabir, A., Al-Olayan, A., 1996. A stocking policy for spare part provisioning under age based preventive replacement. European Journal of Operational Research 90, 171–181.

Kabir, A., Farrash, S., 1996. Simulation of an integrated age replacement and spare provisioning policy using

SLAM. Reliability Engineering and System Safety 52 (2), 129–138.

Law, A., Kelton, W., 1991. Simulation Modeling & Analysis, second ed. McGraw-Hill, New York.

Love, C., Guo, R., 1996. Utilizing weibull failure rates in repair limit analysis for equipment replacement/preventive maintenance decisions. Journal of the Operations Research Society 47 (11), 1366–1376.

Park, Y., Park, S., 1986. Generalized spare ordering policies with random lead time. European Journal of Operational Research 23, 320–330.

Pinedo, M., Chao, X., 1999. Operations Scheduling: With Applications in Manufacturing and Services. McGraw-Hill, New York.

Sleptchenko, A., Van Der Heijden, M., Van Harten, A., 2002. Effects of finite repair capacity in multi-echelon, multi-indenture service part supply systems. International Journal of Production Economics 79, 209–230.

Sleptchenko, A., Van der Heijden, M., Van Harten, A., 2005. Using repair priorities to reduce stock investment in spare part networks. European Journal of Operational Research 163 (3), 733–750.

Tijms, H., 1994. Stochastic Models: An Algorithmic Approach. Wiley, New York.

Van der Heijden, M., Van Harten, A., Ebben, M., 2001. Waiting times at periodically switched one-way traffic lanes. Probability in the Engineering and Informational Sciences 15 (4), 495–518.

Wang, H., 2002. A survey of maintenance policies of deteriorating systems. European Journal of Operational Research 139, 469–489.