ELSEVIER

Reliability Engineering and System Safety 66 (1999) 13–27

RELIABILITY ENGINEERING & SYSTEM SAFETY

www.elsevier.com/locate/ress

Preventive maintenance and the interval availability distribution of an unreliable production system

G. van Dijkhuizen*, M. van der Heijden

School of Management Studies, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received 28 August 1998; accepted 11 January 1999

Abstract

Traditionally, the optimal preventive maintenance interval for an unreliable production system has been determined by maximizing its limiting availability. Nowadays, it is widely recognized that this performance measure does not always provide relevant information for practical purposes. This is particularly true for order-driven manufacturing systems, in which due date performance has become a more important, and even a competitive factor. Under these circumstances, the so-called interval availability distribution is often seen as a more appropriate performance measure. Surprisingly enough, the relation between preventive maintenance and interval availability has received little attention in the existing literature. In this article, a series of mathematical models and optimization techniques is presented, with which the optimal preventive maintenance interval can be determined from an interval availability point of view, rather than from a limiting availability perspective. Computational results for a class of representative test problems indicate that significant improvements of up to 30% in the guaranteed interval availability can be obtained, by increasing preventive maintenance frequencies somewhere between 10 and 70%.

Keywords: Interval availability distribution; Preventive maintenance; Order-driven manufacturing systems

1. Introduction

In studying the performance of an unreliable production system, the limiting availability does not always provide the most relevant information for practical purposes. For example, the amount of gas to be delivered over a finite period of time is often contractually guaranteed in the oil industry [1]. Although short interruptions of the production process can usually be covered by inventory backups, a loss of production for several consecutive days might cause problems in meeting the sales contract, involve high penalty costs, and—in the worst case—loss of goodwill or even customers [25]. In computer and manufacturing systems, the guaranteed performance during a finite period of time is sometimes a more important and even competitive factor than the average performance observed over an infinite horizon [11]. In this respect, the *interval availability* of a production system is often seen as an appropriate performance measure in a practical context. This is particularly true for order-driven manufacturing systems, in which capacity planning plays a key strategical role in satisfying contractual obligations.

Most capacity planning tools used in industry account for random outages by computing average capacity in terms of limiting availability. By doing so, it is immediately clear that during a given period of time (e.g. a week), capacity problems will occur frequently. As this is generally not acceptable, a safety margin is usually built in, in order to ensure satisfactory capacity in e.g. at least 95% of all cases. However, even if this works well in practice, it underlines the point that thinking in terms of the guaranteed capacity of a production system during a finite period of time, is often more appropriate than thinking about its average capacity in the long run. In this respect, a production system with frequent, predictable and short interruptions is to be preferred to one with infrequent, unpredictable and long interruptions, all other things being equal (see Fig. 1). This is a potentially valuable insight, as random breakdowns are one of the major sources of variability.

During the last decades, this and other factors have resulted in an increased popularity of mathematical models for reliability and maintenance optimization, e.g. see [4,10,13,14,16,18,22] for extensive literature reviews. Usually, these models focus on the optimization of system performance in the long run. Frequently encountered approaches include the maximization of the long run system availability, and the minimization of the average

* Corresponding author. Fax: + 31-53-489-2159.
E-mail address: g.c.vandijkhuizen@sms.utwente.nl (G.v. Dijkhuizen)

0951-8320/99/$ - see front matter
PII: S0951-8320(99)00013-7

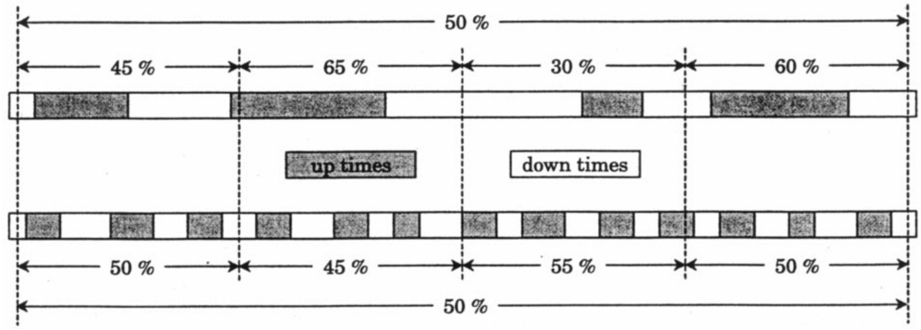

Fig. 1. Production systems with similar limiting availabilities, but different interval availability distributions.

(discounted) total costs per unit of time. Relevant cost factors include the costs of preventive and corrective maintenance, e.g. as a result of inspections, replacements, repairs and /or production losses. At the same time, a growing interest can be observed in modelling the short term behavior of production systems, in terms of the so-called interval availability distribution. The reader is referred to Smith [19] for a comprehensive and up-to-date survey on existing literature. Surprisingly enough, the interactions between preventive maintenance on the one hand, and interval availability on the other hand, have received little attention in existing literature, possibly because of the inherent mathematical complications.

If a production system is repaired at failure, and thus all maintenance is corrective, consecutive up (life) and down (repair) times are usually modelled as stochastically independent random variables. Obviously, this modelling assumption cannot be sustained if preventive maintenance is carried out at regular intervals. In that case, consecutive up and downtimes become mutually dependent random variables, as small up times (caused by failures) are usually followed by large downtimes (caused by repairs), and vice versa. Obviously, this phenomenon does not make life easier from a mathematical point of view. But even for a two-state production system without preventive maintenance, i.e. with alternating and mutually independent up and downtimes, closed-form solutions for the interval availability distribution are not available. Pioneering work on this subject was carried out by Takács [20], who derived an analytical expression consisting of an infinite summations of terms, each consisting of multiple convolutions of the life and repair time distributions. As then, several authors have tried to find reasonable approximations, as well as lower and upper bounds for the interval availability distribution, e.g. see [6–9,12,17,19,23–26].

A three-state production system that is maintained according to an age replacement strategy, is a special case of the class of repairable systems whose state space can be partitioned into a set of up-states and a set of down-states. Csenki [6,7] has shown that these systems can be modelled as semi-Markov chains, for which the cumulative uptime distribution (and hence the interval availability distribution) can be obtained by numerically solving a system of integral equations. His general framework has the advantage of high flexibility to model a wide variety of repairable systems. A drawback of his approach, however, is that the approximation error may be significant. Further, the numerical method requires a lot of computational effort, even considerably more than simulation [7].

In this article, we focus on the optimization of preventive maintenance for a three-state production system. Our main objective and contribution is to provide insight into the effect of short term availability fluctuations on the optimal preventive maintenance strategy. Up to our knowledge, this phenomenon has not been analyzed before in literature. Maintenance optimization means that we have to evaluate a wide range of cumulative uptime distributions for various preventive maintenance strategies. Hence, a numerically efficient method is very important. Therefore, we present a dedicated numerical method for a three-state system, although we are aware of the fact that a more general—but computationally expensive—framework is available [7]. In Section 4.5, we will relate our modelling approach to this more general framework.

2. General approach

Consider an unreliable production system which is repaired upon failure, and maintained preventively as soon as $\theta > 0$ time units have elapsed since the last maintenance action, either preventive or corrective. After preventive and/or corrective maintenance, the system can be considered as good as new. The time to failure or lifetime L of the production system is described by a cumulative distribution function $F(\cdot)$, with probability density function $f(\cdot)$, and corresponding mean $\mu_L > 0$ and variance $\sigma_L^2 \geq 0$. Moreover, the preventive maintenance time P is described by a cumulative distribution function $G(\cdot)$, with mean $\mu_P > 0$ and a variance $\sigma_P^2 \geq 0$. Finally, the corrective maintenance (repair) time R is described by a cumulative distribution function $H(\cdot)$, with mean $\mu_R > 0$ and variance $\sigma_R^2 \geq 0$. As in most maintenance optimization models, we assume that both L, P and R are mutually independent random variables.

2.1. Limiting availability

The limiting availability A_∞ is defined as the fraction of time that the production system is operational (up), if observed over an infinite period of time. If we denote with T_{up} a continuous period of time during which the system is operational, and with T_{down} a continuous period of time during which the system is not operational, then it follows from renewal theory [5] that the limiting availability A_∞ is determined as:

$$A_\infty = \frac{E\{T_{\text{up}}\}}{E\{T_{\text{up}}\} + E\{T_{\text{down}}\}}$$

Depending on the length of the preventive maintenance interval $\theta > 0$, the following expressions can be derived for $E\{T_{\text{up}}\}$ and $E\{T_{\text{down}}\}$. Here, we denote $\bar{F}(\theta) = 1 - F(\theta)$ for notational convencience:

$$E\{T_{\text{up}}\} = \theta\bar{F}(\theta) + \int_0^\theta \tau f(\tau)d\tau$$

$$E\{T_{\text{down}}\} = \mu_P\bar{F}(\theta) + \mu_R F(\theta)$$

If $\theta \to \infty$, this yields $E\{T_{\text{up}}\} = \mu_L$, $E\{T_{\text{down}}\} = \mu_R$, and thus $A_\infty = \mu_L/(\mu_L + \mu_R)$. Traditionally, the optimal preventive maintenance for a production system has been determined in view of maximizing its limiting availability. Unfortunately, this performance measure does not always provide sufficient and relevant information for practical purposes. Sometimes, the so-called interval availability is seen as a more appropriate performance measure.

2.2. Interval availability

The interval availability is defined as the fraction of time that a production system is operational during a given time interval of finite length. Of course, it depends on the initial state of the system at the beginning of this interval, which type of behavior will be observed. From now on, we will assume—without loss of generality—that the production system starts as new at time $t = 0$. If we denote with U_t the cumulative uptime during the interval [0,t], then the interval availability A_t during this interval is defined as follows:

$$A_t = \frac{U_t}{t}.$$

With $T_u = \inf\{t|U_t \geq u\}$, we denote the time required to attain a cumulative uptime of u time units. As both U_t and T_u are random variables, we are mainly interested in their cumulative distribution functions $P(U_t \leq u)$ and $P(T_u \leq t)$. By observing that $P(U_t \leq u) = P(T_u \geq t)$, it is sufficient to determine either $P(U_t \leq u)$ or $P(T_u \leq t)$. Nevertheless, it is known from previous studies into interval availability [15,7] that $P(T_u \leq t)$ is to be preferred from a theoretical point of view, as the corresponding analytical expressions are mathematically more tractable (see also Section 3).

To avoid confusion, we will refer to $P(U_t \leq u)$ as the interval availability distribution, and to $P(T_u \leq t)$ as the *availability interval distribution*. Simply stated, $P(T_u \leq t)$ reflects the probability of completing a cumulative workload of u time units within t units of calendar time. Nowadays, this performance measure could be of considerable interest in e.g. due date determination and order acceptation, as it may provide useful information about the probability that a certain amount of workload will be completed within a certain amount of time.

As an illustrative example, consider a customer order of 10 h processing time which must be completed within 3 days. Moreover, suppose that already 50 h of workload have been accepted for other customers, with their own due dates as well. In case of a first-in-first-out (FIFO) service discipline, which is completely natural in such a setting, this would imply that on-time delivery of this new customer order can be realized with probability $P(T_{50+10} \leq 3 \times 24) = P(T_{60} \leq 72)$. Of course, it is up to management to decide whether or not this is acceptable. Nevertheless, our model could provide useful decision support in this respect. Moreover, it could also be used to explore the opportunities for, and consequences of changing priorities between customer orders.

2.3. Outline

As a starting point, we investigate the initial behavior of the system in Section 3. To be specific, an analytical expression is derived for the probability $P_0(T_u \leq t)$ of at least u units of cumulative uptime during the interval [0,t]. Subsequently, we investigate the limiting behavior of the production system in Section 4, by deriving an analytical expression for the probability $P_\infty(T_u \leq t)$ of at least u units of cumulative uptime during an arbitrary interval of length $t > 0$ in a stabilized situation. In Section 5, some explicit formulas are derived for a production system with Gamma distributed repair and fixed maintenance times, and a simple but efficient algorithm is presented with which the optimal maintenance interval can be determined to a sufficient level of detail. Subsequently, a series of numerical experiments is presented in Sections 6 and 7. Computational results indicate that significant improvements can be obtained in practice, if the optimal preventive maintenance interval is determined from an interval availability rather than a limiting availability point of view. Finally, Section 8 summarizes some conclusions, and identifies some opportunities for further research.

3. Initial behavior of the system

As a starting point, we consider the case where the production system starts with an uptime at time $t = 0$, and preventive maintenance is carried out as soon as the system has been operational (up) for $\theta > 0$ time units. In this respect, a clear distinction must be made between the case $u > \theta$, and the case $u \leq \theta$, as the latter requires much simpler modelling techniques.

3.1. Model without preventive maintenance ($u \leq \theta$)

If the required cumulative uptime $u \leq \theta$, it is immediately clear that no preventive maintenance actions will be involved. As a consequence, all maintenance (if any) will be corrective, and our analysis becomes similar to the well-known failure-based model [20]. As life times L and repair times R are stochastically independent random variables, with corresponding cumulative distribution functions $F(\cdot)$ and $H(\cdot)$, this yields the following expression for $P_0(T_u \leq t)$:

$$P_0(T_u \leq t) = \sum_{n=0}^{\infty} H_n(t-u)\{F_n(u) - F_{n+1}(u)\}$$

Here, $F_n(\cdot)$ and $H_n(\cdot)$ denote the n-fold Stieltjes convolutions of $F(\cdot)$ and $H(\cdot)$ respectively, i.e. $F_1(u) = F(u)$ and $F_n(u) = \int_0^u f(v)F_{n-1}(u-v)\mathrm{d}v$. More specifically, $F_n(u) - F_{n+1}(u)$ denotes the probability of exactly n failures during the first u units of cumulative uptime, whereas the accumulated downtime of these failures does not exceed the amount of $t-u$ time units with probability $H_n(t-u)$. Obviously, this analysis cannot be sustained as soon as subsequent up and downtimes become mutually dependent random variables. Of course, this happens if preventive maintenance is carried out at regular intervals. In that case, the correlation between consecutive up and downtimes usually drops below zero, as typically small (corrective) uptimes go together with large (corrective) downtimes, and large (preventive) uptimes go together with small (preventive) downtimes.

3.2. Model with preventive maintenance ($u > \theta$)

If the required cumulative uptime satisfies $u > \theta$, our analysis proceeds as follows. First of all, we determine the probability $\xi_u^0(m,n)$ of exactly m preventive maintenance actions and n corrective maintenance actions during the first u units of cumulative uptime. Obviously, not all values of m and n correspond to non-zero probabilities $\xi_u^0(m,n)$. As a starting point, as each preventive maintenance action corresponds to exactly θ units of uptime, the number of preventive maintenance actions m should at least satisfy $m\theta < u$. Moreover, as each corrective maintenance action corresponds to at most θ units of uptime, the number of preventive and corrective maintenance actions $m+n$ should also satisfy $(m+n+1)\theta \geq u$.

In all other cases, it is possible to derive an expression for $\xi_u^0(m,n)$. By the complete randomness of consecutive maintenance actions, i.e. preventive with probability $\bar{F}(\theta)$ and corrective with probability $F(\theta)$, the probability $\xi_u^0(m,n)$ must be equal to

$$\binom{m+n}{m}$$

times the probability that exactly m consecutive preventive maintenance actions are followed by exactly n consecutive corrective maintenance actions within the first u units of cumulative uptime. If we denote with $\tilde{F}(t) = P(L \leq t | L \leq \theta) = \min\{1, F(t)/F(\theta)\}$ the conditional cumulative lifetime distribution function, this yields the following expression for $\xi_u^0(m,n)$. Here, $\tilde{F}_n(\cdot)$ denotes the n-fold Stieltjes convolution of $\tilde{F}(\cdot)$, i.e. $\tilde{F}_1(x) = \tilde{F}(x)$, and $\tilde{F}_n(x) = \int_0^x \tilde{f}(y)\tilde{F}_{n-1}(x-y)\mathrm{d}y$ for all $n > 1$:

$$\xi_u^0(m,n) = \binom{m+n}{m}\bar{F}(\theta)^m F(\theta)^n$$
$$\times\{\tilde{F}_n(u-m\theta) - F(\theta)\tilde{F}_{n+1}(u-m\theta)$$
$$- \bar{F}(\theta)\tilde{F}_n(u-(m+1)\theta)\}.$$

The first term between curly brackets reflects the probability that the first m preventive and n corrective maintenance actions are completed within the first u units of cumulative uptime. Similarly, the second and third term reflect the probability that the next i.e. $(m+n+1)$st maintenance action, which is preventive with probability $\bar{F}(\theta)$ and corrective with probability $F(\theta)$, is also completed within the remaining uptime. Together, these terms denote the probability that the first u units of cumulative uptime are attained somewhere between the $(m+n)$st and the $(m+n+1)$st maintenance action, provided that the first m maintenance actions are preventive and the following n maintenance actions are corrective. For notational convenience, and without loss of generality, we will use the notation of $\xi_u^0(m,n)$ in deriving analytical expressions for $P_0(T_u \leq t)$ in the sequel.

3.3. Stochastic repair and stochastic maintenance times

Given the number of preventive maintenance actions m and corrective maintenance actions n, observed with probability $\xi_u^0(m,n)$, the corresponding downtimes do not accumulate to more than $t-u$ time units with probability $G_m \circ H_n(t-u)$. Here, $G \circ H(x) = \int_0^x g(y)H(x-y)\mathrm{d}y$ denotes the well-known Stieltjes convolution for computing the sum of independent stochastic variables. Summarizing, this yields the following expression for $P_0(T_u \leq t)$:

$$P_0(T_u \leq t) = \sum_{m\theta < u \leq (m+n+1)\theta} \xi_u^0(m,n) G_m \circ H_n(t-u).$$

Following a similar argument, the first and higher moments of T_u can be derived in a rather straightforward manner, as long as the corresponding moments of the maintenance and repair time distributions are available. For example, the first two moments of $T_u - u$, given that the system starts with an

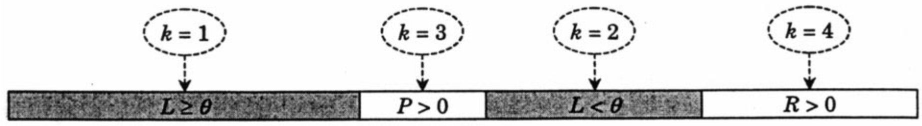

Fig. 2. Possible states of the production system: preventive uptime ($k = 1$), corrective uptime ($k = 2$), preventive downtime ($k = 3$) and corrective downtime ($k = 4$).

uptime at time $t = 0$, are determined as follows:

$$E_0\{T_u - u\} = \sum_{m\theta < u \le (m+n+1)\theta} \xi_u^0(m,n)\{m\mu_P + n\mu_R\}$$

$$E_0\{(T_u - u)^2\} = \sum_{m\theta < u \le (m+n+1)\theta} \xi_u^0(m,n)\{m\sigma_P^2 + n\sigma_R^2 + (m\mu_P + n\mu_R)^2\}.$$

From the transparancy of these expressions, it is immediately clear that the calculation of $P_0(T_u \le t)$ for different values of t, is to be preferred above the calculation of $P_0(U_t \le u)$ for different values of u, at least in view of the inherent mathematical complications.

3.4. Stochastic repair and deterministic maintenance times

If the time required for preventive maintenance is fixed $(\sigma_P = 0)$, the number of preventive maintenance actions m must satisfy $m\mu_P \le t - u$. In that case, $G_m \circ H_n(t-u) = H_n(t - u - m\mu_P)$, and thus an alternative expression can be derived for $P_0(T_u \le t)$:

$$P_0(T_u \le t) = \sum_{m\theta < u \le (m+n+1)\theta, m\mu_p \le t-u} \xi_u^0(m,n) H_n(t - u - m\mu_P).$$

If no failures occur during the first u units of cumulative uptime, the number of preventive maintenance actions equals $\lceil u/\theta - 1 \rceil$ with probability one. Moreover, we know for sure that the corresponding downtimes accumulate to $\lceil u/\theta - 1 \rceil \mu_P$ time units. As each maintenance action is preventive with probability $\bar{F}(\theta)$ and corrective with probability $F(\theta)$, this yields the following additional and a priori information with respect to the cumulative distribution function $P_0(T_u \le t)$:

$$P_0(T_u = u + \lceil u/\theta - 1 \rceil \mu_P) = \bar{F}(\theta)^{\lceil u/\theta - 1 \rceil} \bar{F}(u - \lceil u/\theta - 1 \rceil \theta)$$

A typical example of deterministic maintenance times, but stochastic repair times, can be found in the replacement of single components and/or complete (sub) systems. In general, preventive replacements require a fixed amount of time, as they are perfectly plannable. Corrective replacements, however, often require an additional waiting time, as the required resources are not always readily available on request.

4. Limiting behavior of the system

In this section, we consider the case where the production system starts in an arbitrary state at time $t = 0$, and preventive maintenance is carried out as soon as the system has been operational (up) for $\theta > 0$ time units. As a starting point of our analysis, we observe that the following situations can occur (see Fig. 2):

1. the system starts in a preventive uptime,
2. the system starts in a corrective uptime,
3. the system starts in a preventive downtime,
4. the system starts in a corrective downtime.

Here, a preventive (corrective) uptime is defined as a continuous period of time during which the system is operational (up), and which is *terminated* by a preventive (corrective) maintenance action. This distinction will appear to be convenient when deriving expressions for the availability interval distribution. Similarly, a preventive (corrective) downtime is defined as a continuous period of time during which the system is not operational (down), and which is *initiated* by a preventive (corrective) maintenance action. If we denote with $\tilde{\mu}_L = \int_0^\theta 1 - \tilde{F}(\tau)\mathrm{d}\tau$ the mean length of a corrective uptime, it is easily verified that the corresponding limiting probabilities π_1, π_2, π_3 and π_4 are interrelated as follows. Here, π_k denotes the long run average fraction of time that the system remains in state k:

$$\pi_1 : \pi_2 : \pi_3 : \pi_4 = \theta\overline{F}(\theta) : \tilde{\mu}_L F(\theta) : \mu_P \overline{F}(\theta) : \mu_R F(\theta)$$

Together with the normalization condition $\sum_k \pi_k = 1$, this yields the required and unique values for π_k $(1 \le k \le 4)$. Let us now denote with $P_\mathrm{k}(T_\mathrm{u} \le t)$ the probability of at least u units of cumulative uptime during an arbitrary interval of t time units, given that the system starts in state k. Then obviously, the probability $P_\infty(T_\mathrm{u} \le t)$ of at least u units of cumulative uptime during an arbitrary interval of t time units, given that the system starts in a stationary state, is given by:

$$P_\infty(T_u \le t) = \sum_{k=1}^{4} \pi_k P_k(T_u \le t)$$

In a similar way, the first two moments $E_\infty\{T_\mathrm{u} - u\}$ and $E_\infty\{(T_\mathrm{u} - u)^2\}$ of $T_\mathrm{u} - u$ can be determined. In the following Sections 4.1 to 4.4, we will elaborate on these expressions in more detail. In Section 4.5, we will show that the production system under consideration can also be modelled as a semi-Markov process.

4.1. Start in preventive uptime ($k = 1$)

If the system starts in state $k = 1$, the remaining preventive uptime is described by a random variable $R_1 \in [0, \theta]$ with cumulative distribution function $\Phi_1(\cdot)$:

$$\Phi_1(\tau) = P(R_1 \leq \tau) = \frac{\tau}{\theta}, \qquad 0 \leq \tau \leq \theta.$$

As a starting point, let us determine the probabilities $\xi_u^1(m, n)$ of exactly m preventive and n corrective maintenance actions during the first u units of cumulative uptime, given that the system starts in a preventive uptime ($k = 1$). First of all, we observe that no maintenance occurs if the remaining uptime exceeds the amount of u time units. In all other cases, the first maintenance action must be preventive by assumption. In formula, this yields $\xi_u^1(0, 0) = 1 - \Phi_1(u)$, and $\xi_u^1(0, n) = 0$ for all $n \geq 1$. For all other values of $m \geq 1$ and $n \geq 0$, $\xi_u^1(m, n)$ yields an expression which is similar to $\xi_u^0(m, n)$:

$$\xi_u^1(m, n|m \geq 1) = \binom{m + n - 1}{m - 1} \overline{F}(\theta)^{m-1} F(\theta)^n$$
$$\times \{\Phi_1 \circ \tilde{F}_n(u - (m - 1)\theta - F(\theta)\Phi_1 \circ \tilde{F}_{n+1}(u - (m - 1)\theta)$$
$$- \overline{F}(\theta)\Phi_1 \circ \tilde{F}_n(u - m\theta)\}.$$

Given the number of preventive maintenance actions m and corrective maintenance actions n, the corresponding downtimes do not accumulate to more than $t - u$ time units with probability $G_m \circ H_n(t - u)$. As a consequence, we arrive at the following expression for $P_1(T_u \leq t)$:

$$P_1(T_u \leq t) = \sum_{m,n} \xi_u^1(m, n) G_m \circ H_n(t - u)$$

4.2. Start in corrective uptime ($k = 2$)

If the system starts in state $k = 2$, the remaining corrective uptime is described by a random variable $R_2 \in [0, \theta]$ with cumulative distribution function $\Phi_2(\cdot)$:

$$\Phi_2(\tau) = P(R_2 \leq \tau) = \frac{1}{\bar{\mu}_L} \int_0^\tau 1 - \tilde{F}(v)\mathrm{d}v, \qquad 0 \leq \tau \leq \theta.$$

In a similar way, we can determine the probabilities $\xi_u^2(m, n)$ of exactly m preventive and n corrective maintenance actions during the first u units of cumulative uptime, given that the system starts in a preventive uptime ($k = 2$). As the first maintenance action must be corrective by assumption, we find $\xi_u^2(0, 0) = 1 - \Phi_2(u)$, and $\xi_u^2(m, 0) = 0$ for all $m \geq 1$. For all other values of $m \geq 0$ and $n \geq 1$, $\xi_u^2(m, n)$ yields an expression which is similar to $\xi_u^0(m, n)$ and $\xi_u^1(m, n)$:

$$\xi_u^2(m, n|n \geq 1) = \binom{m + n - 1}{n - 1} \bar{F}(\theta)^m F(\theta)^{n-1}$$
$$\times \{\Phi_2 \circ \tilde{F}_{n-1}(u - m\theta) - F(\theta)\Phi_2 \circ \tilde{F}_n(u - m\theta)$$
$$- \bar{F}(\theta)\Phi_2 \circ \tilde{F}_{n-1}(u - (m + 1)\theta)\}\}$$

Given the number of preventive maintenance actions m and corrective maintenance actions n, the corresponding downtimes do not accumulate to more than $t - u$ time units with probability $G_m \circ H_n(t - u)$. This yields the following expression for $P_2(T_u \leq t)$:

$$P_2(T_u \leq t) = \sum_{m,n} \xi_u^2(m, n) G_m \circ H_n(t - u).$$

Of course, our analysis leads to similar expressions for $P_1(T_u \leq t)$ and $P_2(T_u \leq t)$, as their only difference orginates from the remaining uptimes, with corresponding distribution functions $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$. Note that the combinations of $P_1(T_u \leq t)$ and $P_2(T_u \leq t)$ would not simplify these expressions. In fact, the distinction between preventive and corrective uptimes certainly facilitated their derivation.

4.3. Start in preventive downtime ($k = 3$)

If the system starts in state $k = 3$, the remaining preventive downtime is described by a random variable $R_3 \in [0, \infty\rangle$ with cumulative distribution function $\Phi_3(\cdot)$:

$$\Phi_3(\tau) = P(R_3 \leq \tau) = \frac{1}{\mu_P} \int_0^\tau 1 - G(v)\mathrm{d}v, \qquad \tau \geq 0.$$

As a starting point, we observe that the first uptime starts as soon as preventive maintenance is completed. Therefore $\xi_n^0(m, n)$ denotes the probability of exactly m preventive and n corrective maintenance actions during the first u units of cumulative uptime. Given the number of preventive maintenance actions m and corrective maintenance actions n, the corresponding downtimes do not accumulate to more than $t - u$ time units with probability $\Phi_3 \circ G_m \circ H_n(t - u)$. This yields the following expression for $P_3(T_u \leq t)$:

$$P_3(T_u \leq t) = \sum_{m,n} \xi_u^0(m, n)\Phi_3 \circ G_m \circ H_n(t - u).$$

4.4. Start in corrective downtime ($k = 4$)

If the system starts in state $k = 4$, the remaining corrective downtime is described by a random variable $R_4 \in [0, \infty\rangle$ with cumulative distribution function $\Phi_4(\cdot)$:

$$\Phi_4(\tau) = P(R_4 \leq \tau) = \frac{1}{\mu_R} \int_0^\tau 1 - H(v)\mathrm{d}v, \qquad \tau \geq 0$$

In a similar way, we observe that the first uptime starts as soon as corrective maintenance is completed. Given the number of preventive maintenance actions m and corrective maintenance actions n, with probability $\xi_u^0(m, n)$, the corresponding downtimes do not accumulate to more than $t - u$ time units with probability $\Phi_4 \circ G_m \circ H_n(t-u)$. This yields the following expression for $P_4(T_u \le t)$:

$$P_4(T_u \le t) = \sum_{m,n} \xi_u^0(m,n)\Phi_4 \circ G_m \circ H_n(t-u)$$

Once again, these expressions for $P_3(T_u \le t)$ and $P_4(T_u \le t)$ are very similar, as their only difference originates from remaining down times, with corresponding distribution functions $\Phi_3(\cdot)$ and $\Phi_4(\cdot)$.

4.5. An equivalent semi-Markov model

As mentioned in the introduction, the system can also be modelled as a semi-Markov process [7]. To this end, we define three system states, namely the up-state ($k = 1$), preventive maintenance ($k = 2$) and corrective maintenance ($k = 3$). Note that there is no need to distinguish between preventive and corrective up times now. The semi-Markov process is defined by the kernel matrix $Q(t) = \{q_{kl}(t)\}$, where $q_{kl}(t) = r_{kl}F_{kl}(t)$. Here, r_{kl} represents the one-step transition probability from state k to state l. Moreover, $F_{kl}(t)$ denotes the cumulative distribution function of the holding time in state k, given that a transition to state l is made next. The specification of r_{kl} and $F_{kl}(t)$ for the production system under consideration is given below. Here, $I_S(t)$ reflects the indicator function, i.e. $I_S(t) = 1$ if $t \in S$, and $I_S(t) = 0$ otherwise.

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \begin{pmatrix} 0 & 1 - F(\theta) & F(\theta) \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} F_{11}(t) & F_{12}(t) & F_{13}(t) \\ F_{21}(t) & F_{22}(t) & F_{23}(t) \\ F_{31}(t) & F_{32}(t) & F_{33}(t) \end{pmatrix} = \begin{pmatrix} 0 & I_{[\theta,\infty\rangle}(t) & F(\min\{t,\theta\})/F(\theta) \\ G(t) & 0 & 0 \\ H(t) & 0 & 0 \end{pmatrix}.$$

As a next step, the system states are partitioned into a set of up states $U = \{1\}$ and a set of down states $D = \{2, 3\}$. In the terminology of Csenki [7], our interest goes out to the function $P_k\{T_u \le t\} = WMA_k(u, t-u)$, where $WMA_k(t_1, t_2)$ denotes the so-called Work Mission Availability. The latter function is defined as the probability that no more than t_2 units of time are spent in the set of down states D for a mission of length t_1, given that the system just entered state k at time $t = 0$. The Work Mission Availability for a given set of initial states S is given by the vector $\mathbf{WMA}_S(t_1, t_2)$. These vector values can be obtained by solving the following set of integral equations numerically.

$$\mathbf{WMA}_U(t_1,t_2) = (1 - F(t_1))I_{[0,\theta\rangle}(t_1) + \int_0^{t_1}\int_0^{t_2} \mathbf{J}_{UU}(w_1,w_2)\mathbf{WMA}_U(t_1 - w_1, t_2 - w_2)\mathrm{d}w_1\,\mathrm{d}w_2$$

$$\mathbf{WMA}_D(t_1,t_2) = \begin{pmatrix} G(t_2)(1 - F(t_1))I_{[\theta,\infty\rangle}(t_1) \\ H(t_2)(1 - F(t_1))I_{[\theta,\infty\rangle}(t_1) \end{pmatrix} + \int_0^{t_1}\int_0^{t_2} \mathbf{J}_{DD}(w_1,w_2)\mathbf{WMA}_D(t_1 - w_1, t_2 - w_2)\mathrm{d}w_1\,\mathrm{d}w_2$$

Here, the matrix functions $\mathbf{J}_{UU}$ and $\mathbf{J}_{DD}$ are defined as:

$$\mathbf{J}_{UU}(t_1,t_2) = G(t_2)(1 - F(\theta)I_{[\theta,\infty\rangle}(t_1) + H(t_2)F(\min\{t_1,\theta\})$$

$$\mathbf{J}_{DD}(t_1,t_2) = \begin{pmatrix} G(t_2)(1 - F(\theta))I_{[\theta,\infty\rangle}(t_1) & G(t_2)F(\min\{t_1,\theta\}) \\ H(t_2)(1 - F(\theta))I_{[\theta,\infty\rangle}(t_1) & H(t_2)F(\min\{t_1,\theta\}) \end{pmatrix}$$

This implies that two systems of one resp. two integral equations should be solved numerically, compared to our dedicated approach which requires numerical evaluation of single expressions only. From a computational point of view, the latter is to be preferred in this specific situation, where we aim for optimization of the preventive maintenance interval. In Sections 5 and 6, we will show that our numerical experiments clearly confirm this hypothesis.

5. The optimal maintenance interval

So far, we have considered the preventive maintenance interval θ to be a given constant. In this section, we will present a rather straightforward algorithm with which an optimal preventive maintenance interval θ^* can be determined from an interval availability point of view. To this end, a plausible choice for an objective function is presented first. Subsequently, the objective function under consideration is evaluated for a production system with Gamma distributed repair and fixed maintenance times. In that case, explicit formulas can be derived, which strongly reduce the complexity of the optimization problem.

5.1. Objective functions

In classical maintenance theory, an optimal preventive maintenance interval $\theta_0 < \infty$ for a production system is usually determined by maximizing its limiting availability A_∞ (see Section 2.1). In our setting here, a similar approach would be to minimize the expected time $E_\infty\{T_u\}$ required to attain a cumulative uptime of u time units, given that the system starts in an arbitrary state at time $t = 0$. In general, these objectives are not equivalent, i.e. $E_\infty\{T_u\}.A_\infty \neq u$, in particular if u is relatively small compared to the expected lifettime of the production system. Anyhow, none of these objectives accounts for the fact that $Var\{T_u\}$ is also a value of great interest, as it provides information about the short term behavior of the production system. As an alternative, we have chosen to minimize the *ω-percentile* of the availability interval T_u, where $0 < \omega < 1$ is a user-defined constant. This is quite natural, as it provides information about the one-sided confidence interval for the required time to complete a cumulative workload of u time units. In line with this, our objective becomes to minimize $f_\omega(\theta|u)$, with parameters u and ω:

$$f_\omega(\theta|u) = \inf\{t \geq u | P_\infty(T_u \leq t|\theta) \geq \omega\}.$$

5.2. Function evaluation

Unfortunately, the evaluation of $f_\omega(\theta|u)$ for a given value of θ is rather complicated from a mathematical point of view. As a starting point, numerical approximations for the convolutions $\tilde{F}_n, \Phi_1\circ\tilde{F}_n$, and $\Phi_2\circ\tilde{F}_n$ have to be calculated, in order to determine $\xi_u^0(m,n)$, $\xi_u^1(m,n)$ and $\xi_u^2(m,n)$ for all $m,n \geq 0$. In this respect, an upper bound M respectively N on the number of preventive respectively corrective maintenance actions m respectively n has to be identified, in order to truncate the infinite summations appearing in $P_k(T_u \leq t)$. For $k = 0$, this can be done in a rather straightforward manner, if one realizes that the following relation holds for all $M, N \geq 0$:

$$P_0(T_u \leq t) - \sum_{m=0}^{M}\sum_{n=0}^{N} \xi_u^0(m,n) G_m\circ H_n(t-u) \leq 1 - \sum_{m=0}^{M}\sum_{n=0}^{N} \xi_u^0(m,n).$$

In other words, M and N can be increased in a stepwise manner, until this restriction is satisfied. Obviously, similar results can be obtained for other values of k, which yields the desired result. Subsequently, the stationary probabilities $\pi_k(1 \leq k \leq 4)$, as well as the convolutions $\Phi_3\circ G_m\circ H_n$ and $\Phi_4\circ G_m\circ H_n$, have to be numerically approximated in order to evaluate $P_\infty(T_u \leq t)$ for a given value of t. Finally, a one-dimensional search procedure (e.g. bi-section) has to be carried out in order to identify the smallest value of t for which $P_\infty(T_u \leq t) \geq \omega$. In general, i.e. for arbitrary distribution functions $F(\cdot), G(\cdot)$ and $H(\cdot)$, this yields a complex procedure which requires a large amount of computational effort. Under some special conditions, however, the complexity of evaluating $f_\omega(\theta|u)$ can be reduced significantly, in particular if repair times are Gamma distributed random variables, and preventive maintenance times are fixed.

5.3. Gamma distributed repair and fixed maintenance times

In this section, we will restrict ourselves to the case where repair times are Gamma distributed random variables with parameters $\alpha = \mu_R^2\sigma_R^{-2}$ and $\beta = \mu_R^{-1}\sigma_R^2$, and preventive maintenance requires a fixed amount of time $\mu_P > 0$ (i.e. $\sigma_P = 0$). Under these assumptions, explicit formulas can be derived for $G_m\circ H_n$, $\Phi_3\circ G_m\circ H_n$ and $\Phi_4\circ G_m\circ H_n$, which appear in the definitions $P_k(T_u \leq t)$. As a starting point, we define $\Psi_{\alpha,\beta}(\cdot)$ for notational convenience:

$$\Psi_{\alpha,\beta}(\tau) \equiv \int_0^\tau \Gamma_{\alpha,\beta}(v)dv = \tau\Gamma_{\alpha,\beta}(\tau) - \alpha\beta\Gamma_{\alpha+1,\beta}(\tau)$$

Lemma 1. *If repair times are Gamma distributed random variables with parameters α and β, i.e.* $\mu_R = \alpha\beta$ and $\sigma_R^2 = \alpha\beta^2$, *and preventive maintenance requires a fixed amount of time* $\mu_P > 0$, *then* $G_m\circ H_n(t-u), \Phi_3\circ G_m\circ H_n(t-u)$ and $\Phi_4\circ G_m\circ H_n(t-u)$ *can be derived analytically by means of the following explicit formulas*:

$$G_m\circ H_n(t-u) = \Gamma_{n\alpha,\beta}(t-u-m\mu_p)$$

$$\Phi_3\circ G_m\circ H_n(t-u) = \frac{\Psi_{n\alpha,\beta}(t-u-m\mu_P) - \Psi_{n\alpha,\beta}(t-u-(m+1)\mu_P)}{\mu_P}$$

$$\Phi_4\circ G_m\circ H_n(t-u) = \frac{\Psi_{n\alpha,\beta}(t-u-m\mu_P) - \Psi_{(n+1)\alpha,\beta}(t-u-m\mu_P)}{\mu_R}$$

For a proof of this lemma, we refer to Appendix A. Here, we only mention that efficient computer programming codes are available for the calculation of Gamma distributions, e.g. see [21]. Hence, the only convolutions that need to be numerically approximated are $\tilde{F}_n$, $\Phi_1\circ\tilde{F}_n$, and $\Phi_2\circ\tilde{F}_n$, which appear in the definitions of $\xi_u^0(m,n)$, $\xi_u^1(m,n)$, and $\xi_u^2(m,n)$. As $\tilde{F}(\theta) = \Phi_1(\theta) = \Phi_2(\theta) = 1$ by definition, and thus $\tilde{F}(\cdot)$, $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ all have finite support, these convolutions can be determined to a sufficient level of detail within reasonable computational times.

In this respect, another but intuitively less attractive possibility, is to model both preventive and corrective maintenance times as Gamma distributed random variables with the same shape parameter β. This would imply, however, that either average preventive maintenance times are larger than average corrective maintenance times ($\mu_P > \mu_R$), or the coefficient of variation of preventive maintenance times is larger than the coefficient of variation of corrective maintenance times ($\sigma_P/\mu_P > \sigma_R/\mu_R$). As none of these alternatives is likely to occur in practice, these assumptions would

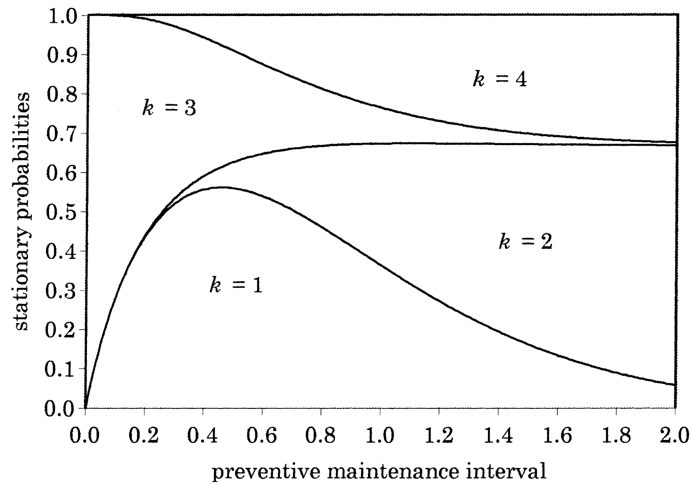

Fig. 3. Stationary probabilities π_k in relation to the preventive maintenance interval θ, for a production system with Gamma distributed life times ($\mu_L = 1, \sigma_L = 1/2$), Gamma distributed repair times ($\mu_R = 1/2, \sigma_R = 1/4$) and fixed preventive maintenance times ($\mu_P = 1/4, \sigma_P = 0$).

leave us with a theoretical exercise of almost no practical relevance.

5.4. Optimization algorithm

Our optimization algorithm starts with observing that $f_\omega(\theta|u)$ has an infinite number of discontinuities of the form $\theta = u/k (k \geq 1)$, because the number of preventive maintenance actions equals $\lfloor u/\theta \rfloor$ or $\lfloor u/\theta \rfloor$ if no failures occur during the required interval of u units cumulative uptime. Therefore, and to avoid the risk of sub-optimization, we decomposed our global optimization procedure into a series of consecutive local optimization procedures within disjunct ranges of the form $[u/(k+1), u/k\rangle$. As a starting point, however, we determine the optimal maintenance interval $\theta_0 < \infty$ from a limiting availability point of view [2], and assume that $\theta^* \leq \theta_0$. In line with this, our first range under consideration becomes $[u/(k_0, \theta_0\rangle$, where $k_0 = \lceil u/\theta_0 \rceil$.

Our optimization algorithm is now based on the assumption that $f_\omega(\theta|u)$ is a piecewise unimodal function within each of the previously mentioned ranges. As a starting point, we determine the optimal preventive maintenance interval $\theta_{k_0}*$ within the range $[u/k_0, \theta_0\rangle$ with the use of golden section search [3]. In a similar way, and starting with $k = k_0$, we determine the optimal maintenance interval θ_k* within the range $[u/(k+1), u/k\rangle$, and at the same time keep track of the best-so-far maintenance interval θ^* within the range $[u/(k+1), \theta_0\rangle$. As soon as the optimal maintenance interval θ^* does not change in two consecutive iterations, the algorithm is terminated. Obviously, this stop criterion is based on the underlying assumption that $f_\omega(\theta_{k+1}*|u) \geq f_\omega(\theta_k*|u)$ implies that $\theta^* \geq \theta_{k+1}*$.

Under some weak conditions, it can be shown that this procedure would lead to the optimal preventive maintenance interval, if we were concerned with the limiting availability of the production system [2]. Unfortunately, we have not been able to prove this results for the availability interval distribution. However, we have had no indications so far that these assumptions strongly affect the performance of our numerical optimization algorithm. Anyhow, the computational results of the following sections should be interpreted as a lower bound for the savings that can be obtained if the optimal maintenance interval is determined from an interval rather than a limiting availability perspective.

6. Numerical example

Let us now present a numerical example in order to illustrate the previously mentioned methods in more detail. To this end, we consider a production system with Gamma distributed life times ($\mu_L = 1$ and $\sigma_L = 1/2$), Gamma distributed repair times ($\mu_R = 1/2$ and $\sigma_R = 1/4$) and fixed preventive maintenance times ($\mu_P = 1/4$ and $\sigma_P = 0$). Moreover, we assume that the required cumulative uptime equals $u = 1$ time units, which is exactly equal to the expected life time of the production system.

6.1. Stationary probabilities

As a starting point, we determine the stationary probabilities π_k of starting in state k, where $1 \leq k \leq 4$ (see Fig. 3). As can be seen from this figure, π_3 tends to one as θ tends to zero. In that case, the production system is maintained preventively all the time, and the system is always down for preventive maintenance. For similar reasons, both π_1 and π_3 tend to zero if θ tends to infinity. In that case, all maintenance will be corrective, and the system will be either up or down, with corresponding probabilities $\pi_2 : \pi_4 = \mu_L : \mu_R = 2 : 1$. Obviously, this corresponds to a limiting availability of $A_\infty = \mu_L/(\mu_L + \mu_R) = 2/3$. From Fig. 3, it

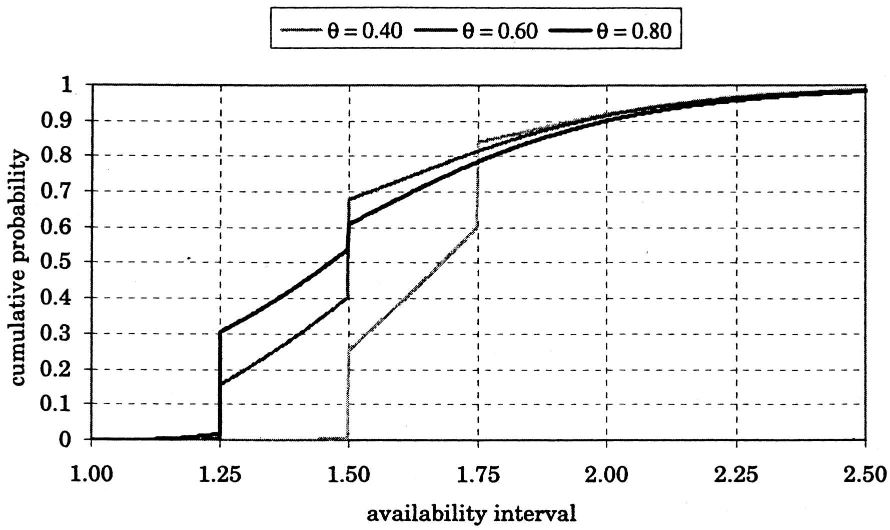

Fig. 4. Availability interval distribution $P_\infty(T_u \leq t)$ in case u = 1, for a production system with Gamma distributed life times ($\mu_L = 1, \sigma_L = 1/2$), Gamma distributed repair times ($\mu_R = 1/2, \sigma_R = 1/4$), fixed preventive maintenance times ($\mu_P = 1/4, \sigma_P = 0$), and different preventive maintenance intervals θ.

can also be concluded that the optimal maintenance interval $\theta_0 \approx 1.10$ if our objective is to maximize the limiting availability $A_\infty = \pi_1 + \pi_2$ of the production system. Moreover, this optimal maintenance interval hardly outperforms a corrective maintenance strategy ($\theta^* \to \infty$).

6.2. Availability interval distribution

To continue our analysis, let us determine the limiting behavior of the production system for different values of θ, in terms of the cumulative distribution function $P_\infty(T_u \leq t)$. In this particular example, we determined these probabilities for $\theta = 4/5$, $\theta = 3/5$, and $\theta = 2/5$, respectively (see Fig. 4). Along the horizontal axis, the discontinuities $u + \lfloor u/\theta \rfloor \mu_P$ and $u + \lceil u/\theta \rceil \mu_P$ for T_u are clearly visible. Moreover, there are some strong indications that the optimal maintenance interval is closely related to the desired confidence interval. To be specific, the best maintenance interval equals $\theta = 4/5$ for $\omega = 1/2$, $\theta = 3/5$ for $\omega = 3/4$, and $\theta = 2/5$ for $\omega = 7/8$. Apparently, the optimal maintenance interval decreases if the guaranteed performance during a finite period of time (interval availability) becomes more important than the average performance during an infinite period of time (limiting availability).

6.3. Optimal maintenance interval

In order to arrive at the optimal maintenance interval θ^*, let us now determine $f_\omega(\theta | u = 1)$ for different values of both $\omega \in \{0.90, 0.95, 0.99\}$ and $\theta \in \langle 0, 2]$. The results are

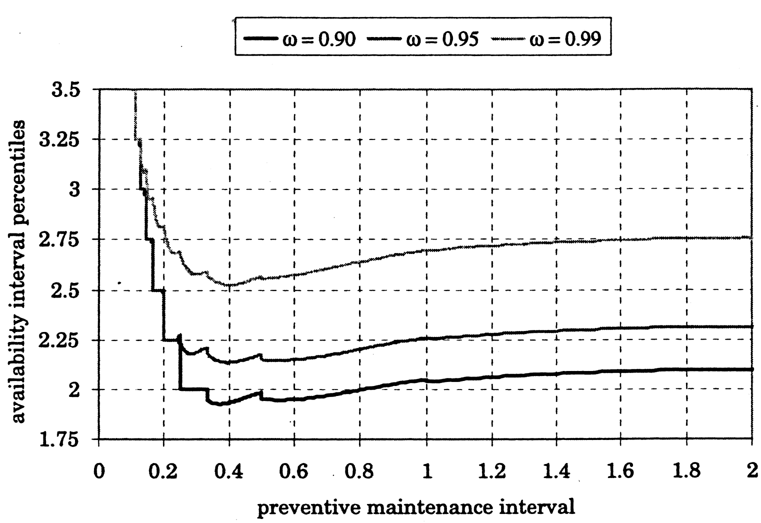

Fig. 5. Percentiles of the availability interval T_u in case $u = 1$, in relation to the preventive maintenance interval θ, for a production system with Gamma distributed life times ($\mu_L = 1, \sigma_L = 1/2$), Gamma distributed repair times ($\mu_R = 1/2, \sigma_R = 1/4$) and fixed preventive maintenance times ($\mu_P = 1/4, \sigma_P = 0$).

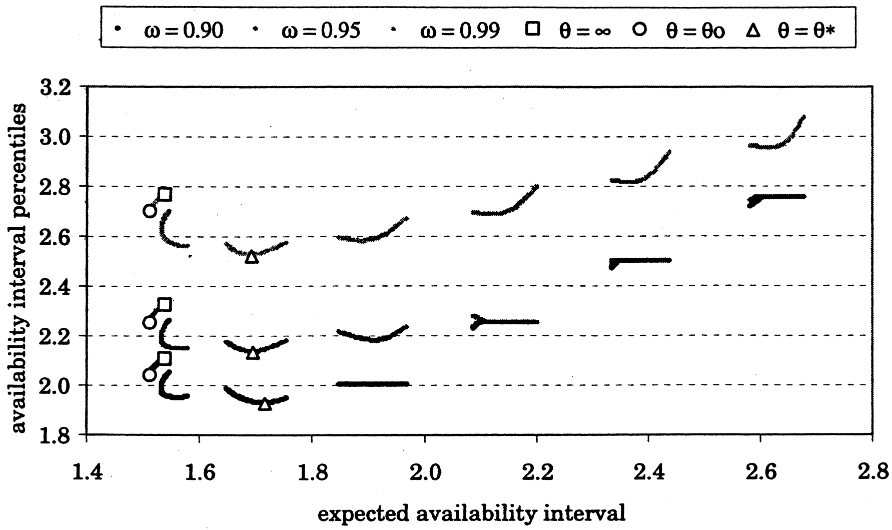

Fig. 6. Expectation $E_\infty\{T_u\}$ versus ω-percentiles $f_\omega(\theta|u)$ of the availability interval T_u , in case $u = 1$, for a production system with Gamma distributed life times ($\mu_L = 1, \sigma_L = 1/2$), Gamma distributed repair times ($\mu_R = 1/2, \sigma_R = 1/4$), fixed preventive maintenance times ($\mu_P = 1/4, \sigma_P = 0$), and different maintenance intervals θ.

depicted in Fig. 5. A closer look at these results provided the following optimal maintenance intervals: $\theta^* = 0.38$ for $\omega = 0.90$, $\theta^* = 0.40$ for $\omega = 0.95$ and $\theta^* = 0.40$ for $\omega = 0.99$. Apparently, the optimal maintenance interval θ^* hardly depends on the value of ω in this particular example. Moreover, we can conclude from Fig. 5 that the optimal maintenance interval $\theta^* \approx 0.40$ is significantly smaller than the optimal maintenance interval $\theta_0 \approx 1.10$ from a limiting availability point of view. In the following section, we will carry out a series of numerical experiments in order to investigate the relation between θ^* and θ_0 on the one hand, and between $f_\omega(\theta^*|u)$ and $f_\omega(\theta_0|u)$ on the other hand.

6.4. Limiting versus interval availability

Let us now further elaborate upon the difference between the average and guaranteed performance of the production system. To this end, we compared the expectation $E_\infty\{T_1\}$ and the ω-percentiles $f_\omega(\theta|u = 1) = \inf\{t \geq 1|P_\infty\{T_1 \leq t\} \geq \omega\}$ of the time T_1 required to attain a cumulative uptime of exactly $u = 1$ time units. Of course, this was done for different values of θ and $\omega \in \{0.90, 0.95, 0.99\}$. The results are depicted in Fig. 6. As we expected, the discontinuities of the form $\theta = u/k$ are clearly visible, and cause empty spaces in this figure. Moreover, we observe that a corrective maintenance strategy ($\theta \to \infty$) performs poor in both dimensions. Starting from here, decreasing the maintenance interval leads to an improvement in both dimensions, up to the point where the expected value $E_\infty\{T_u\}$ is minimized ($\theta \approx \theta_0$). Subsequently, reducing the maintenance interval leads to degradations in the first dimension, but at the same time to further improvements in the second dimension, up to the point where the ω-percentile $f_\omega(\theta|u)$ is minimized ($\theta = \theta^*$). At this point, further reductions in the preventive maintenance interval leads to degradations in both dimensions.

6.5. Computational effort

We conclude this section by briefly discussing the computational effort that was required to arrive at a reasonable estimate of the ω-percentile of T_u, for a specific value of θ. In this particular example, i.e. with $u = 1$, it took us no more than 10 s on a personal computer (80486) to approximate the ω-percentile of T_u for an arbitrary value of θ. Moreover, an extensive simulation study showed that these approximations were already very accurate, with relative errors of less than 1‰. Although one should be aware of the fact that the number of preventive and/or corrective maintenance actions—and thus computation times—tend to increase (linearly) with u, these observations give rise to the attractibility of our dedicated numerical method in relation to a more general semi-Markov modelling framework (see [7] for details).

7. Computational results

In this section, we will present the results of a series of numerical experiments that were carried out for a production system with Gamma distributed lifetimes, Gamma distributed repair times, and fixed preventive maintenance times. Amongst other factors, the main objectives of these numerical experiments were (i) to determine what happens if the optimal maintenance interval is determined from an interval availability rather than a limiting availability point of view, and (ii) to investigate how the optimal maintenance intervals for interval availability depends on the characteristics of the production system.

For notational convenience, and without loss of generality,

Table 1
Comparison of the optimal maintenance intervals θ^* for interval availability and θ_0 for limiting availability, as well as the corresponding ω-percentiles $f_\omega(\theta^*)$ and $f_\omega(\theta_0)$, for different values of u and α

		$\frac{\theta_0 - \theta^*}{\theta_0} \times 100\%$			$\frac{f_\omega(\theta_0) - f_\omega(\theta^*)}{f_\omega(\theta_0)} \times 100\%$		
		Minimal	Average	Maximal	Minimal	Average	Maximal
$\omega = 0.90$	$u = 1$	20.6	41.2	67.6	0.0	6.7	17.8
	$u = 2$	18.3	29.6	45.4	0.7	3.4	9.4
	$u = 3$	19.5	26.1	35.9	0.4	2.4	6.7
$\omega = 0.95$	$u = 1$	20.6	39.7	64.6	1.9	10.9	20.8
	$u = 2$	10.6	32.6	46.4	0.8	5.5	14.7
	$u = 3$	19.5	29.9	37.6	0.6	3.4	9.7
$\omega = 0.99$	$u = 1$	35.6	49.1	63.7	2.3	12.1	29.0
	$u = 2$	32.8	43.4	49.9	1.3	7.4	21.0
	$u = 3$	18.4	36.7	46.5	0.7	4.7	13.6

we assumed that $\mu_L = 1$ in each test problem. Moreover, the relevant parameters μ_R/μ_L, μ_P/μ_R, σ_L/μ_L and σ_R/μ_R were varied between 1/2 and 1/4 in order to arrive at 16 production systems with—at least theoretically—different short term behavior. For each production system, we generated a total of 9 test problems by choosing $u \in \{1, 2, 3\}$ and $\omega \in \{0.90, 0.95, 0.99\}$. For each of the 144 test problems obtained this way, the optimal maintenance interval θ^* for interval availability, the optimal maintenance interval θ_0 for limiting availability, and the corresponding availability interval percentiles $f_\omega(\theta^*)$ and $f_\omega(\theta_0)$ were determined. An overview of all test problems is depicted in Table 1. Moreover, the results for all test problems with $u = 1$ are depicted in Table 2.

As a starting point, it is easily verified from Table 1 that significant improvements can be obtained in the short term behavior of a production system, if the optimal maintenance interval is determined from an interval availability rather than a limiting availability point of view. Depending on the required amount of cumulative uptime $u > 0$, the required percentile $\omega < 1$, and the characteristics of the production system, the corresponding improvements are substantial, with a maximum of about 30% in the availability interval. To achieve this, a 10% to 70% reduction in the preventive maintenance interval was typical.

It is intuitively clear, and can easily be derived as well, that the optimal maintenance interval θ^* converges to θ_0 if the required uptime u tends to infinity. In general, the required uptime $u > 0$ and percentile $\omega < 1$ on the one hand, and the optimal maintenance intervals θ^* and θ_0 with availability interval percentiles $f_\omega(\theta^*)$ and $f_\omega(\theta_0)$ on the other hand, are interrelated as follows:

- an increase in the desired uptime u usually goes together with an increase in the optimal maintenance interval θ^* for interval availability, as well as a decrease in the relative performance of θ^* versus θ_0 in terms of $f_\omega(\theta_0)/f_\omega(\theta^*)$;
- an increase in the desired percentile ω usually goes together with a decrease in the optimal maintenance

Table 2
Availability interval percentiles for T_u at the optimal maintenance intervals θ^* for interval availability and θ_0 for limiting availability, as observed in 16 test problems with $u = 1$ and $\mu_L = 1$

$\frac{\mu_R}{\mu_L}$	$\frac{\mu_P}{\mu_R}$	$\frac{\sigma_L}{\mu_L}$	$\frac{\sigma_R}{\mu_R}$	θ_0	$\omega = 0.90$ θ^*	$f(\theta^*)$	$f(\theta_0)$	$\omega = 0.95$ θ^*	$f(\theta^*)$	$f(\theta_0)$	$\omega = 0.99$ θ^*	$f(\theta^*)$	$f(\theta_0)$
0.5	0.5	0.5	0.5	1.10	0.37	1.93	2.05	0.40	2.13	2.27	0.40	2.52	2.71
0.5	0.5	0.5	0.25	1.10	0.64	1.89	1.96	0.64	2.03	2.11	0.43	2.30	2.39
0.5	0.5	0.25	0.5	0.78	0.50	1.50	1.71	0.50	1.50	1.89	0.50	1.74	2.26
0.5	0.5	0.25	0.25	0.78	0.50	1.50	1.70	0.50	1.50	1.81	0.34	1.75	2.01
0.5	0.25	0.5	0.5	0.56	0.26	1.50	1.82	0.20	1.71	2.03	0.22	2.23	2.48
0.5	0.25	0.5	0.25	0.56	0.26	1.50	1.76	0.37	1.84	1.91	0.28	2.07	2.25
0.5	0.25	0.25	0.5	0.63	0.50	1.25	1.25	0.50	1.25	1.54	0.40	1.38	1.94
0.5	0.25	0.25	0.25	0.63	0.50	1.25	1.25	0.50	1.25	1.56	0.40	1.38	1.77
0.25	0.5	0.5	0.5	1.10	0.36	1.46	1.51	0.39	1.56	1.62	0.40	1.76	1.84
0.25	0.5	0.5	0.25	1.10	0.62	1.44	1.48	0.62	1.51	1.55	0.43	1.65	1.69
0.25	0.5	0.25	0.25	0.78	0.50	1.25	1.35	0.50	1.25	1.44	0.50	1.36	1.62
0.25	0.5	0.25	0.5	0.78	0.50	1.25	1.35	0.50	1.25	1.40	0.50	1.37	1.51
0.25	0.25	0.5	0.5	0.56	0.26	1.25	1.40	0.20	1.32	1.50	0.21	1.61	1.72
0.25	0.25	0.5	0.25	0.56	0.26	1.25	1.37	0.37	1.42	1.44	0.28	1.53	1.62
0.25	0.25	0.25	0.5	0.63	0.50	1.13	1.13	0.50	1.13	1.26	0.39	1.19	1.46
0.25	0.25	0.25	0.25	0.63	0.50	1.13	1.13	0.50	1.13	1.28	0.39	1.19	1.39

interval θ^* for interval availability, as well as an increase in the relative performance of θ^* versus θ_0 in terms of $f_\omega(\theta_0)/f_\omega(\theta^*)$.

In a similar way, it can be derived from Table 2 to which extent the characteristics of the production system affect the optimal maintenance intervals θ^*, as well as the corresponding availability interval percentiles $f_\omega(\theta^*)$. As a starting point, and as expected, we observe that θ_0 depends on σ_L/μ_L and μ_P/μ_R only. In addition, the following observations were made from Table 2:

- an increase in the ratio of repair versus maintenance times usually goes together with a decrease in the optimal maintenance interval θ^* for interval availability, as well as a decrease in the corresponding availability interval percentile $f_\omega(\theta^*)$;
- an increase in the variation of life and/or repair times usually goes together with a decrease in the optimal maintenance interval θ^* for interval availability, as well as an increase in the corresponding availability interval percentile $f_\omega(\theta^*)$.

Although intuitively attractive, these rules of thumbs do not cover all possible situations that may occur, e.g. see $\mu_R/\mu_L = \mu_P/\mu_R = 0.5$ and $\sigma_L/\mu_L = \sigma_R/\mu_R = 0.25$ in Table 2. Nevertheless, we conclude that the guaranteed availability interval of a production system can be improved significantly, if the optimal preventive maintenance interval is determined from an interval availability perspective. From a practical point of view, this means that preventive maintenance is a powerful instrument to increase the *controllability* or *predictability* of a production system.

8. Concluding remarks

In this article, we have presented a series of mathematical models which can be used to determine the availability interval distribution for a production system which is maintained preventively at regular intervals, according to an age replacement strategy. Moreover, we have presented an optimization algorithm, with which the optimal maintenance interval can be determined from an availability interval point of view. A series of numerical experiments indicated that significant improvements in the availability interval can be obtained in comparison with a classical limiting availability perspective, and that these effects become stronger as the variabilities in life and/or repair times increase. Simply stated, our computational results have illustrated that preventive maintenance does not only increase the availability, but also reduces the variability of a production system, and that the latter is often a more important performance measure. Although these conclusions were drawn within a setting of Gamma distributed repair and fixed maintenance times, we strongly believe that they are also applicable to more complex systems.

To conclude this chapter, let us now briefly discuss the possibilities for approximating the availability interval distribution, in case repair times are not Gamma distributed random variables and/or preventive maintenance times are not fixed. In the most general case, it not possible to derive explicit formulas for the convolutions $G_m \circ H_n$ appearing in $P_k(T_u \le t)$. Under such circumstances, another interesting and potentially promising approach is to fit a Gamma (or other) distribution to the first two moments $E_\infty\{T_u - u\}$ and $E_\infty\{(T_u - u)^2\}$ of $T_u - u$. The underlying observation behind this approach is that these moments can be determined as long as the first three moments of $G(\cdot)$ and $H(\cdot)$ are available. In general, we may have to account for the fact that the availability interval distribution $P_\infty(T_u \le t)$ might have some discontinuities as well. If these jumps are known in advance, i.e. in terms of a set $\Omega = \{t \ge u | P_\infty(T_u = t) > 0\}$ of availability intervals with non-zero probabilities, it seems worthwile to approximate $P_\infty(T_u \le t | t \notin \Omega)$ with the use of $E_\infty\{T_u - u | t \notin \Omega\}$ and $E_\infty\{(T_u - u)^2 | t \notin \Omega\}$. These suggestions, however, are left for future research.

Acknowledgements

The authors would like to thank the referees for pointing us at additional relevant literature, and for their useful suggestions to improve the first version of our article. In particular, we would like to thank Attila Csenki (University of Bradford) for reminding us that the production system under consideration could also be analysed with the use of semi-Markov models.

Appendix A. Proof of Lemma 1

As a starting point, we observe that the following expressions can be derived for the cumulative distribution functions $\Phi_3(\cdot)$ and $\Phi_4(\cdot)$ of the remaining preventive downtime $R_3 \in [0, \mu_P]$, and the remaining corrective downtime $R_4 \in [0, \infty\rangle$:

$$\Phi_3(\tau) = \frac{\tau}{\mu_P}, \qquad 0 \le \tau \le \mu_P$$

$$\Phi_4(\tau) = \frac{1}{\mu_R} \int_0^\tau 1 - \Gamma_{\alpha,\beta}(v)\mathrm{d}v, \qquad \tau \ge 0$$

For notational convenience, and without loss of generality, we substitute $z = t - u - m\mu_P$ in the sequel. Our analysis now proceeds as follows. First of all, we observe that the expressions for $G_m \circ H_n(t-u)$ and $\Phi_3 \circ G_m \circ H_n(t-u)$ can be derived rather straightforwardly by some elementary algebra. In contrast, $\Phi_4 \circ G_m \circ H_n(t-u)$ yields a somewhat more complex expression, which must be further simplified:

$$G_m \circ H_n(t-u) = H_n(z) = \Gamma_{n\alpha,\beta}(z)$$

$$\Phi_3 \circ G_m \circ H_n(t-u) = \Phi_3 \circ \Gamma_{n\alpha,\beta}(z)$$

$$= \frac{1}{\mu_P}\int_0^{\mu_P} \Gamma_{n\alpha,\beta}(z-\tau)\mathrm{d}\tau$$

$$= \frac{\Psi_{n\alpha,\beta}(z) - \Psi_{n\alpha,\beta}(z-\mu_P)}{\mu_P}$$

$$\Phi_4 \circ G_m \circ H_n(t-u) = \Phi_4 \circ \Gamma_{n\alpha,\beta}(z)$$

$$= \frac{1}{\mu_R}\int_0^{z} (1-\Gamma_{\alpha,\beta}(\tau))\Gamma_{n\alpha,\beta}(z-\tau)\mathrm{d}\tau$$

$$= \frac{\Psi_{n\alpha,\beta}(z) - \int_0^z \Gamma_{\alpha,\beta}(\tau)\Gamma_{n\alpha,\beta}(z-\tau)\mathrm{d}\tau}{\mu_R}$$

Apparently, we need to show that $\Psi_{(n+1)\alpha,\beta}(z) = \int_0^z \Gamma_{\alpha,\beta}(\tau)\Gamma_{n\alpha,\beta}(z-\tau)\mathrm{d}\tau$ in order to arrive at the expression for $\Phi_4 \circ G_m \circ H_n(t-u)$ in Lemma 1. Our analysis now proceeds as follows. First of all, we observe that $\int_0^z \Gamma_{\alpha,\beta}(\tau)\Gamma_{n\alpha,\beta}(z-\tau)d\tau$ can be rewritten as follows:

$$\int_0^z \Gamma_{\alpha,\beta}(\tau)\Gamma_{n\alpha,\beta}(z-\tau)\mathrm{d}\tau = \int_0^z\int_0^\tau \frac{\mathrm{d}}{\mathrm{d}v}\{\Gamma_{\alpha,\beta}(v)\Gamma_{n\alpha,\beta}(z-v)\}\mathrm{d}v\,\mathrm{d}\tau$$

$$= \int_0^z\int_0^\tau \gamma_{\alpha,\beta}(v)\Gamma_{n\alpha,\beta}(z-v)\mathrm{d}v\,\mathrm{d}\tau - \int_0^z\int_0^\tau \Gamma_{\alpha,\beta}(v)\gamma_{n\alpha,\beta}(z-v)\mathrm{d}v\,\mathrm{d}\tau$$

By changing the integration variables, both integrals can be reduced to one-dimensional integrals, each of which can be evaluated explicitly by using the following well-known properties $\int_0^z \tau\gamma_{\alpha,\beta}(\tau)\mathrm{d}\tau = \alpha\beta\Gamma_{\alpha+1,\beta}(z)$ and $\int_0^z \gamma_{\alpha_1,\beta}(\tau)\Gamma_{\alpha_2,\beta}(z-\tau)d\tau = \Gamma_{\alpha_1+\alpha_2,\beta}(z)$ for Gamma distributions:

$$\int_0^z\int_0^\tau \gamma_{\alpha,\beta}(v)\Gamma_{n\alpha,\beta}(z-v)\mathrm{d}v\,\mathrm{d}\tau = \int_0^z (z-v)\gamma_{\alpha,\beta}(v)\Gamma_{n\alpha,\beta}(z-v)\mathrm{d}v$$

$$= z\int_0^z \gamma_{\alpha,\beta}(v)\Gamma_{n\alpha,\beta}(z-v)dv - \int_0^z v\gamma_{\alpha,\beta}(v)\Gamma_{n\alpha,\beta}(z-v)\mathrm{d}v$$

$$= z\Gamma_{(n+1)\alpha,\beta}(z) - \alpha\beta\Gamma_{(n+1)\alpha+1,\beta}(z)$$

$$\int_0^z\int_0^\tau \Gamma_{\alpha,\beta}(v)\gamma_{n\alpha,\beta}(z-v)\mathrm{d}v\,\mathrm{d}\tau = \int_0^z (z-v)\Gamma_{\alpha,\beta}(v)\gamma_{n\alpha,\beta}(z-v)\mathrm{d}v$$

$$= \int_0^z v\gamma_{n\alpha,\beta}(v)\Gamma_{\alpha,\beta}(z-v)\mathrm{d}v = n\alpha\beta\Gamma_{(n+1)\alpha+1,\beta}(z)$$

As $z\Gamma_{(n+1)\alpha,\beta}(z) - (n+1)\alpha\beta\Gamma_{(n+1)\alpha+1,\beta}(z) = \Psi_{(n+1)\alpha,\beta}(z)$, this completes the proof.

References

[1] Aven T. On performance measures for multistate monotone systems. Reliability Engineering and System Safety 1993;41:259–266.

[2] Barlow R, Proschan F. Mathematical Theory of Reliability. New York: Wiley, 1965.

[3] Brent R. Algorithms for Minimization without Derivatives. Englewood Cliffs: Prentice Hall, 1973.

[4] Cho D, Parlar M. A survey of maintenance models for multi-unit systems. European Journal of Operational Research 1991;51:1–23.

[5] Cox D. Renewal Theory. New York: Wiley, 1962.

[6] Csenki A. Transient analysis of interval availability for repairable systems modelled by semi-markov processes: IMA Journal of Mathematics Applied in Business and Industry 1995;6:267–281.

[7] Csenki A. A new approach to the cumulative operational time for semi-markov models of repairable systems. Reliability Engineering and System Safety 1996;54:11–21.

[8] De Souza e Silva E, Gail H. Calculating cumulative operational time distributions of repairable computer systems. IEEE Transactions on Computers 1986;35:322–332.

[9] De Souza e Silva E, Gail H. Calculating availability and performability measures of repairable computer system using randomization. Journal of the Association for Computing Machinery 1989;36:171–193.

[10] Dekker R, Wildeman R, Van der Duyn Schouten F. A review of multicomponent maintenance models with economic dependence. Mathematical Methods of Operations Research 1997;45:411–435.

[11] Goyal A, Tantawi A. A measure if guaranteed availability and its numerical evaluation. IEEE Transactions on Computers 1988;37:25–32.

[12] Haukaas H, Aven T. Formulae for the down time distribution of a monotone system observed in a time interval. Reliability Engineering and System Safety 1996;52:19–26.

[13] McCall J. Maintenance policies for stochastically failing equipment: a survey. Management Science 1965;11:493–524.

[14] Ozekici S, editor. Reliability and Maintenance of Complex Systems Berlin: Springer, 1996.

[15] Para A, Garribba S. Reliability analysis of a repairable single unit under general age-dependence. Synthesis and Analysis Methods for Safety and Reliability Studies, Proceedings of the NATO Advanced Study Institute, Urbino, Italy, 3–14 July 1978, New York: Plenum Press, 1980. pp. 251–268.

[16] Pierskalla W, Voelker J. A survey of maintenance models: the control and surveillance of deteriorating systems. Naval Research Logistics Quarterly 1976;23:353–388.

[17] Schäbe H. Computing instationary availability for maintained systems. Microelectronics Reliability 1996;36:55–70.

[18] Sherif Y, Smith M. Optimal maintenance models for systems subject to failure—a review. Naval Research Logistics Quarterly 1981;28:47–74.

[19] Smith M. An approximation of the interval availability distribution.

Probability in the Engineering and Informational Sciences 1997;11:451–467.

[20] Takács L. On certain sojourn time problems in the theory of stochastic processes. Acta Mathematica Academiae Scientiarum Hungaricae 1957;8:169–191.

[21] Temme N. A set of algorithms for the incomplete Gamma functions. Probability in the Engineering and Informational Sciences 1994;8:291–307.

[22] Valdez-Flores C, Fledman R. A survey of preventive maintenance models for stochastically deteriorating single-unit systems. Naval Research Logistics 1989;36:419–446.

[23] Van der Heijden M. Interval availability distribution for a 1-out-of-2 reliability system with repair. Probability in the Engineering and Informational Sciences 1987;1:211–224.

[24] Van der Heijden M, Schornagel A. Interval uneffectiveness distribution for a k-out-of-n multistate reliability system with repair. European Journal of Operational Research 1988;36:66–77.

[25] Van Rijn C, Schornagel A. STAMP, a new technique to assess the probability distribution of system effectiveness (Proceedings of the 17th annual meeting of the System Reliability Service). The SRS Quarterly digest 1987;17:50–58.

[26] Wartenhorst P, Performance Analysis of Repairable Systems, Ph.D. thesis, Katholieke Universiteit Brabant, 1993.