# Maintenance Policies for Non-repairable Components

**Bram de Jonge**

## 1 Introduction

We consider maintenance policies for non-repairable components. We consider a component as a part of a system that is subject to maintenance interventions and for which no further subdivisions are made into sub-components that are individually subject to any maintenance interventions. The condition of a non-repairable component cannot be partially improved by carrying out a repair; maintenance of a non-repairable component is therefore always a replacement. In most cases, such a replacement will result in a component that is as-good-as-new. Only if a heterogeneous set of spare components is considered, the quality of a new components differs per replacement.

A component can be replaced either after its failure or before its failure. In the first case we talk about corrective, reactive, or failure-based maintenance; the second case is referred to as preventive maintenance. It is generally preferred to perform maintenance interventions preventively, for instance because failure of a component can result in damage to other components, and because it can lead to unplanned downtime. However, performing preventive maintenance too often is also undesirable and costly. Therefore, a balance has to be found between the preventive maintenance frequency and the risk of failures.

A maintenance policy describes when to carry out preventive maintenance. A distinction can be made between time-based maintenance policies and condition-based maintenance (CBM) policies. The former is based on the time that a component is in service, the latter allows for maintenance activities that are performed based on degradation information.

B. de Jonge (✉)
Department of Operations, University of Groningen, Groningen, The Netherlands
e-mail: b.de.jonge@rug.nl

Time-based maintenance is easy to implement as only the time that a component is in service has to be recorded. However, substantial remaining useful life is wasted if the machine is still in reasonable condition when preventive maintenance is performed, and a breakdown might occur if it happens to deteriorate faster than expected. Condition-based maintenance, on the other hand, generally results in more effectively scheduled preventive maintenance, and, in the ideal case, preventive maintenance that is performed just before failure. However, applying condition-based maintenance is only possible if there are conditions that are related to the moment of failure, and if it is technically possible to monitor these conditions. Furthermore, condition-based maintenance should only be applied if its benefits outweigh the efforts and costs required to apply it. These requisites include condition monitoring equipment and software to store, analyze, and initiate maintenance actions.

## 2 Time-Based Maintenance

Traditionally, two time-based preventive maintenance policies can be distinguished, viz. age-based maintenance and block-based maintenance (Barlow and Proschan, 1965; Gertsbakh, 2000). Under the age-based maintenance policy, corrective maintenance is performed when the component fails, and preventive maintenance is performed when the age of the component reaches $T$, whichever occurs first (see Fig. 1). The maintenance age $T$ is the decision variable of this policy. Under the block-based maintenance policy (sometimes also called periodic maintenance), preventive maintenance is performed at fixed times $kT$, $k = 1, 2, \ldots$. Corrective maintenance is performed when the component fails, but this does not affect the preventive maintenance schedule (see Fig. 2). The maintenance interval $T$ is the decision variable of this policy. The disadvantage of block-based maintenance is that preventive maintenance is sometimes performed shortly after a failure. The main advantages, on the other hand, are the easier planning as it is known in advance when preventive maintenance will be performed, and the clustered maintenance actions if the same block-based policy is used for multiple components.

We let $F$ denote the (cumulative) distribution function of the time until failure of the component. We will consider time-based maintenance from a cost perspective. The cost of performing a preventive maintenance action is denoted by $c_{\text{pm}}$, the cost
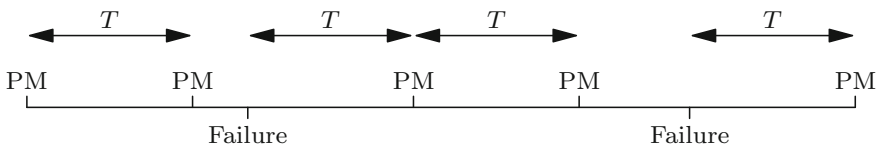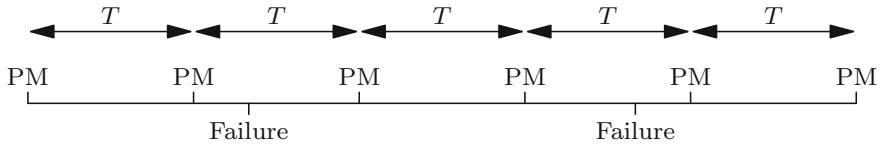


**Fig. 1** Scheme of the age-based maintenance policy

**Fig. 2** Scheme of the block-based maintenance policy

of a corrective maintenance action by $c_{cm}$. The cost of preventive maintenance is assumed to be lower than the cost of corrective maintenance, i.e., $c_{pm} < c_{cm}$, implying that preventive maintenance can be beneficial when scheduled effectively. In the basic models both preventive and corrective maintenance actions are assumed to require a negligible amount of time and to make the component as-good-as-new. The cost of performing corrective maintenance is often normalized to 1, so that only one cost parameter $c$ for the relative cost of performing preventive maintenance is required.

### 2.1 Age-Based Maintenance

The (long-run) cost rate (i.e., the long-run mean cost per unit of time) of the age-based maintenance policy depends on the maintenance age $T$ and is denoted by $\eta_{age}(T)$. Because both types of maintenance make the component as-good-as-new, standard renewal theory can be used to evaluate this cost rate. By referring to the time between consecutive maintenance actions as a cycle, the cost rate can be written as

$$\eta_{age}(T) = \frac{\text{Mean cost per cycle}}{\text{Mean cycle length}}$$
$$= \frac{c_{cm}F(T) + c_{pm}(1 - F(T))}{\int_0^T (1 - F(x))\, dx}.$$

Studies that consider the age-based maintenance policy typically assume that the lifetime distribution is known with certainty. De Jonge et al. (2015) acknowledge that this is often not realistic, and they consider the optimal age-based maintenance policy under uncertainty in the lifetime distribution. They assume a certain parametric lifetime distribution and include uncertainty in its parameters.

In general, they represent the vector of parameters of the lifetime distribution by $s$ and denote the joint density function that models the uncertainty in $s$ by $g(s)$, which is defined on $\mathbb{R}^n$. Instead of the cost rate we can now talk about the expected

cost rate $\eta_{\text{age}}^{E}(T)$ as a function of the maintenance age $T$:

$$\eta_{\text{age}}^{E}(T) = \int_{s \in \mathbb{R}^n} g(s) \frac{c_{\text{cm}} F(T; s) + c_{\text{pm}}(1 - F(T; s))}{\int_0^T (1 - F(x; s)) \, \mathrm{d}x} \, \mathrm{d}s_1 \cdots \mathrm{d}s_n.$$

The preventive maintenance age $T_{\text{opt}}^{E}$ that minimizes this expected cost rate is considered as the optimal maintenance age.

De Jonge et al. (2015) start to consider a uniform lifetime distribution with uncertainty in its right end point; this uncertainty is modeled by a second uniform distribution. Although the uniform distribution is not the most realistic lifetime distribution, this setting has the advantage that it can be analyzed algebraically.

The authors continue to consider a Weibull lifetime distribution, which is the most commonly used distribution to model lifetimes. The Weibull distribution has a shape parameter $k$ and a scale parameter $\lambda$. Because the failure mode of a component often provides an accurate estimation for the shape parameter $k$, there is in practice generally most uncertainty in the scale parameter $\lambda$. The authors model the uncertainty in $\lambda$ by using a uniform distribution on the interval $[1 - \alpha, 1 + \alpha]$. The value of $\alpha \in [0, 1]$ can be interpreted as a measure for the level of uncertainty in $\lambda$. This setting needs to be analyzed numerically.

Figure 3 shows the optimal maintenance age as a function of the level of uncertainty $\alpha$ in the scale parameter $\lambda$ of a Weibull lifetime distribution with
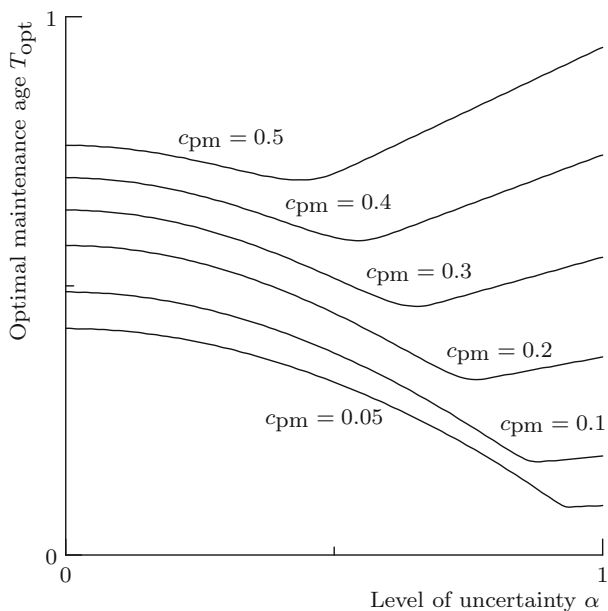


**Fig. 3** Optimal preventive maintenance age under uncertainty in the scale parameter $\lambda$ of a Weibull lifetime distribution with shape parameter $k = 5$, corrective maintenance cost $c_{\text{cm}} = 1$, and for various preventive maintenance costs $c_{\text{pm}}$

$k = 5$. It turns out that the optimal maintenance age first decreases in the level of uncertainty. If the level of uncertainty exceeds a certain threshold the optimal maintenance age starts to increase. The initial decrease is expected; more uncertainty in the lifetime distribution results in earlier preventive maintenance. However, if the uncertainty increases further, it becomes too expensive to prevent very early failures. Longer lifetimes also become more likely when the uncertainty increases; this results in an increasing maintenance age.

A similar pattern is observed when a uniform lifetime distribution with uncertainty in its right end point is considered. This parameter basically also is the scale parameter of this distribution. We also expect a similar pattern when uncertainty in the scale parameter of other parametric lifetime distributions is considered, and when the uncertainty in the scale parameter itself is modeled by a different distribution. We would also like to mention that parametric bootstrapping has also been used to obtain the probability distribution of an estimator for the optimal maintenance age (Tokumoto et al., 2014).

In the setting above a static decision is considered that is not updated when more information becomes available. However, the distribution that models the uncertainty can be updated when more data becomes available. When a failure occurs an event duration is obtained, whereas a preventive maintenance action results in censored durations. Both types of durations can be used to update the uncertainty in a Bayesian manner.

Event durations are more informative than censored durations, and long censored durations are more informative than short censored durations. In other words, the choice of a maintenance age influences the information that becomes available. This is acknowledged by De Jonge et al. (2015); they suggest to postpone preventive maintenance actions at the start of the lifespan of a component. This will result in an increase in costs during the first phase of the lifespan of the components, but it also results in reduced uncertainty and thereby in more effectively scheduled maintenance actions during the remainder of the lifespan. The aim is to find a balance so that the total costs during the entire lifespan is minimized. In the literature, this tradeoff is also referred to as the exploration–exploitation dilemma.

Because (De Jonge et al., 2015) are the first to recognize that the choice of the maintenance policy influences the information that becomes available, they have considered a simple setting with only two component types, viz., weak and strong components. Both component types have a Weibull lifetime distribution with a common value of the shape parameter $k$; the values of the respective scale parameters $\lambda$ are different. The knowledge is modeled by the estimated probability that the component is strong, and a threshold policy is used that postpones preventive maintenance as long as this probability exceeds a certain threshold, i.e., as long as it is not sufficiently sure that the component is weak. This threshold policy is compared to a policy that minimizes the expected cost rate based on the current knowledge as described above. It turns out that the threshold policy can offer substantial cost reductions as opposed to the policy that minimizes the expected cost rate.

The previous analysis is based on a Weibull lifetime distribution with uncertainty only in its scale parameter. Although most uncertainty is often in the scale parameter, there also exist situations in which uncertainty in the shape parameter is expected. This can be the case if the failure mode of equipment is not known, or if there are multiple competing failure modes. This may lead to interesting results because a shape parameter $k < 1$ corresponds to a decreasing failure rate, implying that preventive maintenance is never beneficial. The optimal policy in settings where it is not known whether there is an increasing or a decreasing failure rate is of interest.

Another avenue for future research is to assume that the parametric distribution itself is not known, i.e., to assume model uncertainty instead of parameter uncertainty. A difficulty of such settings is that a selection of candidates for the true parametric distribution has to be made, and that prior probabilities need to be specified. Moreover, other optimality criteria instead of the expected cost rate could be considered. Minimization of the expected cost rate leads to the best decisions on average, but these decisions may be unacceptable for certain values of the unknown parameters.

## 2.2 Block-Based Maintenance

For the block-based maintenance policy the renewal points are the times at which preventive maintenance is performed. Renewal cycles thus always have length $T$, and the preventive maintenance cost is incurred once per cycle (at the end of each cycle). We let $m(t)$ denote the expected number of failures during a period with length $t$ that starts with a component that is as-good-as-new, and during which no preventive maintenance is performed. The cost rate $\eta_{\text{block}}(T)$ of the block-based maintenance policy as a function of the preventive maintenance interval $T$ equals

$$\eta_{\text{block}}(T) = \frac{c_{\text{pm}} + c_{\text{cm}}m(T)}{T}. \tag{1}$$

The main difficulty in evaluating $\eta_{\text{block}}(T)$ is that it requires the evaluation of the mean number of failures $m(T)$ during a time period with length $T$. The function $m(t)$ is called a renewal function and can be calculated as

$$m(t) = \sum_{n=1}^{\infty} F_n(t),$$

in which $F_n$ represents the $n$th convolution of the lifetime distribution function $F$. The first convolution $F_1$ equals the distribution function $F$ itself; the other convolutions can be determined recursively:

$$F_n(t) = \int_0^t f(x) F_{n-1}(t-x) \mathrm{d}x, \quad n = 2, 3, \ldots$$

In practice, $m(t)$ is often approximated numerically by using the first few convolutions. This generally results in good approximations because the number of failures to expect in between consecutive preventive maintenance actions is typically low.

Studies that consider a block-based maintenance policy generally assume that machines or components are either used continuously, or that the deterioration does not depend on the actual usage. In practice, however, this is often not realistic. De Jonge and Jakobsons (2018) consider the block-based maintenance policy for a component that is not used continuously and for which the actual usage is random. Furthermore, the component is assumed to only deteriorate when it is active. Although the future usage is stochastic, it is assumed that all maintenance actions have to be scheduled in advance, and therefore a block-based maintenance policy is considered.

The authors model the random component usage by a Markov switching. The component is alternately active and idle, and the lengths of these periods are modeled by exponential durations. Active periods are exponentially distributed with rate parameter $\alpha_1$, whereas idle periods are exponentially distributed with rate parameter $\alpha_0$. It follows that active periods have mean length $1/\alpha_1$ and that idle periods have mean length $1/\alpha_0$, from which it follows that the usage rate $\rho$ of the component is given by

$$\rho = \frac{\frac{1}{\alpha_1}}{\frac{1}{\alpha_1} + \frac{1}{\alpha_0}} = \frac{\alpha_0}{\alpha_0 + \alpha_1}.$$

As mentioned before, the main difficulty in evaluating the cost rate (1) of the block-based maintenance policy is the evaluation of the renewal function $m(t)$. In the current setting with random component usage there is not even a closed-form expression for the distribution function $F$ of the time until failure. There are, however, two limiting cases that can be analyzed using the renewal function of the lifetime distribution. We will denote this renewal function by $m_W(t)$. If, for instance, the component has a Weibull lifetime distribution, then $m_W$ is the renewal function of the Weibull distribution.

The two limiting cases are those with a very high and with a very low switching frequency. If the switching frequency is very high, the usage in between two preventive maintenance actions is very stable. Approximately, the component will be active during time period $\rho T$ in between two consecutive preventive maintenance actions, and failures can only occur during this time period. Thus, the expected number of failures during the maintenance interval can be approximated by $m_W(\rho T)$, and the cost rate (1) by

$$\eta_{\text{freq}}(T) = \frac{c_{\text{pm}} + c_{\text{cm}} m(T)}{T} \approx \frac{c_{\text{pm}} + c_{\text{cm}} m_W(\rho T)}{T}.$$

In the other limiting case the switching frequency is very low. This implies that, in between two consecutive preventive maintenance actions, it is very likely that the component is either entirely active, or entirely idle. The corresponding probabilities are $\rho$ and $1 - \rho$, respectively, with $\rho$ equal to the usage rate (2). Failures can only occur if the component is active, implying that the expected failure cost during a maintenance interval is $c_{\mathrm{cm}} \rho m_W(T)$. In this case the cost rate can be approximated by

$$\eta_{\mathrm{rare}}(T) = \frac{c_{\mathrm{pm}} + c_{\mathrm{cm}} m(T)}{T} \approx \frac{c_{\mathrm{pm}} + c_{\mathrm{cm}} \rho m_W(T)}{T}.$$

Because the usage is quite stable for high switching frequencies, this limiting case results in a relatively long preventive maintenance interval. In order to avoid failures during long active periods, a much more conservative preventive maintenance interval is optimal for low switching frequencies. De Jonge and Jakobsons (2018) analyze the general case of the problem by formulating it as a set of integral equations. They show that the optimal maintenance interval and the corresponding cost rate for more moderate switching frequencies are in between the two bounds obtained from the two limiting cases. Furthermore, they also show that, for moderate switching frequencies, it is important to choose the maintenance interval based on the actual usage pattern, instead of only based on the usage rate of the component.

Future research in this area could consider active and idle periods that are not exponentially distributed. In such a setting one has to keep track of the time that the component is already active or idle, which complicates the analysis. Instead of analyzing this setting algebraically, it would also be possible to use simulations. Another possibility for future developments could be to consider multiple component speeds, instead of only on and off. This means that more sophisticated stochastic models are needed to model the random usage of the component. Random component usage can also be relevant in settings with condition-based maintenance. In such settings there is often a planning time between initiating and performing preventive maintenance. A component that is not used continuously during the planning time is expected to result in a higher optimal deterioration level at which preventive maintenance is scheduled. Finally, in the above, it is assumed that the component usage is dictated externally. However, if there is some flexibility in the usage, the performance of the system would benefit from the possibility to simultaneously optimize maintenance and usage decisions.

## 3 Condition-Based Maintenance

Because of the increasing possibilities to monitor, store, and analyze condition information of equipment, condition-based maintenance (CBM) policies are gaining popularity. A prerequisite for analyzing and optimizing condition-based maintenance policies is the modeling of deterioration processes of components.

Distinctions between deterioration processes can be made based on the state space (either discrete or continuous), and on the time scale (also either discrete or continuous).

Another important distinction that can be made is that between continuous condition monitoring and condition monitoring based on inspections. The first case is applicable if a sensor is used for condition monitoring; in this case we continuously know what the actual deterioration level of the component is. When inspections are needed to obtain condition information, we do not only need to determine when to carry out maintenance, but we also need to determine an inspection schedule or policy.

Inspection schedules are either periodic or aperiodic. The advantage of periodic inspections is that the entire inspection schedule is fully specified by a single decision variable, namely the time between consecutive inspections. This eases both the optimization and the implementation in a practical industrial context. However, when acceptable in practice, aperiodic inspections are often preferred because failure becomes more likely as the deterioration level increases. A final note is that an entire aperiodic inspection schedule can be fixed in advance, but that the next inspection can also be scheduled dynamically based on the currently observed deterioration level.

## 3.1 Delay-Time Model

The most simple deterioration model is the so-called delay-time model. It is a continuous-time model that adds a "deteriorated" state in between the operating state and the failed state. Thus, the model has three states. It is called the delay-time model because a delayed failure occurs after reaching the deteriorated state. When considering the delay-time model, probability distributions have to be specified for the time until reaching the deteriorated state, and for the time in between reaching the deteriorated state and failure. Most studies that adopt the delay-time model assume that an inspection is required to observe the deteriorated state and that failures are self-announcing. Analysis is easiest if the exponential distribution is used to model the time until reaching the deteriorated state. In that case, if immediate preventive maintenance is carried out when an inspection reveals the deteriorated state, all inspections are renewal points.

Although the delay-time model is proposed by Christer (1976) in 1976, there are new developments in delay-time modeling to date. For instance, (Van Oosterom et al., 2014) consider a periodic inspection schedule, but they relax the common assumption that preventive maintenance should be carried out immediately when an inspection reveals the deteriorated state. Instead, they allow the maintenance action to be delayed. The advantage is twofold. First, the utilization of the useful life of the component is improved, and second, the maintenance cost is reduced as a result of a longer time window to prepare maintenance resources. Wang et al. (2017) allow for a delayed first inspection, and a periodic inspection schedule

thereafter. Furthermore, they initially schedule a replacement at a certain age. If an inspection reveals the deteriorated state and the time until the age-based replacement is less than a certain threshold level, then the preventive replacement action will be delayed. Otherwise, the component will be replaced immediately.

## 3.2 Gamma Deterioration Process with Continuous Monitoring

A commonly used continuous-time continuous-state stochastic deterioration process is the stationary gamma process. The gamma process was introduced in the area of reliability by Abdel-Hameed (1975). It has the property that the deterioration increments, within any time interval of any length, are gamma distributed with identical scale parameter.

The density function $f$ of the gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ equals

$$f_{\alpha,\beta}(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0,$$

in which $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} \, dz$ denotes the gamma function. The stationary gamma process has a shape function $at$ with shape parameter $a > 0$ and a scale parameter $b > 0$. It is a continuous-time process $\{X(t) : t \geq 0\}$ with the following properties:

1. $X(0) = 0$ with probability 1.
2. $X(\tau) - X(t) \sim f_{a(\tau-t),b}$ for $\tau > t \geq 0$.
3. $X(t)$ has independent increments.
4. $X(t)$ is a jump process with infinitely many jumps in any time interval.

The process is stationary because the increments $X(\tau) - X(t)$ depend only on $\tau - t$ for all $\tau$ and $t$. The expectation and the variance of the process $X(t)$ are given by

$$\mathrm{E}(X(t)) = abt \quad \text{and} \quad \mathrm{Var}(X(t)) = ab^2 t,$$

respectively. Thus, the variance of the deterioration process, relative to its mean, is small if $a$ is large and $b$ is small, and is large if $a$ is small and $b$ is large. We will use the standard deviation $\sigma = \sqrt{a} \cdot b$ as a measure for the amount of volatility in the stationary gamma deterioration process. Figure 4 shows sample paths of stationary gamma processes with $\sigma = 0.05$, $\sigma = 0.5$, and $\sigma = 5$.

De Jonge et al. (2017) consider a single maintainable component that is monitored continuously and for which the deterioration is modeled by a stationary gamma process. Failure occurs when the amount of deterioration exceeds a given level $L$. After such a failure an immediate corrective maintenance action will

**Fig. 4** Sample paths of
stationary gamma processes
for various standard
deviations $\sigma$
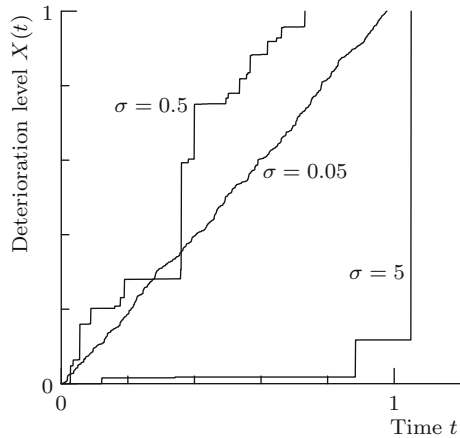


be carried out. Furthermore, as long as the component is functioning, preventive
maintenance can be carried out. The costs of preventive and corrective maintenance
are again denoted by $c_{\mathrm{pm}}$ and $c_{\mathrm{cm}}$, respectively. Both types of maintenance are
assumed to require a negligible amount of time and to make the component as-
good-as-new, i.e., they will bring the deterioration level back to 0.

The aim of the study is to compare the performance of condition-based mainte-
nance to the performance of time-based maintenance. The condition-based mainte-
nance policy is prescribed by a single deterioration threshold level $M$. Preventive
maintenance is performed when the deterioration level exceeds this level $M$. This
commonly used policy is called the control-limit policy. The threshold $M$ should
not be chosen too close to the failure level $L$ because the deterioration process is
a jump process. In other words, when $M$ is close to $L$ and when the deterioration
level exceeds $M$, it may also immediately jump over $L$, resulting in failure. The
time-based maintenance policy that is considered is the age-based maintenance
policy. Thus, preventive maintenance is carried out if a certain maintenance age
$T$ is reached, see also Sect. 2.1.

Figure 5 shows the cost rate of the condition-based maintenance (CBM) policy as
a function of the preventive maintenance threshold $M$, and the cost of the time-based
maintenance (TBM) policy as a function of the maintenance age $T$. The gamma
process is specified by $a = 5$ and $b = 0.22$ (this results in a mean time to failure
of 1), the failure threshold equals $L = 1$, and the cost parameters are $c_{\mathrm{pm}} = 0.2$
and $c_{\mathrm{cm}} = 1$. Simulation has been used to make the figure. It turns out that the cost
rate under the optimal CBM policy is substantially lower than the cost rate under
the optimal TBM policy. In other words, the availability of condition information
results in substantial cost savings. It can also be observed that the optimal preventive
maintenance threshold $M$ is much smaller than the failure threshold $L = 1$. As
explained before, this is caused by the fact that the deterioration process is a jump
process.

**Fig. 5** The cost rate under
the CBM policy (as a function
of the $M$) and under the TBM
policy (as a function of $T$)
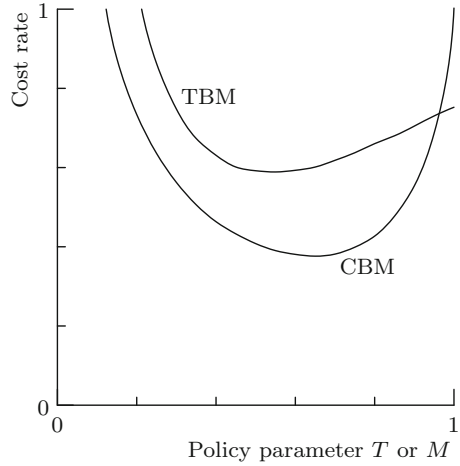


**Fig. 6** The cost rate under
the optimal CBM and the
optimal TBM policy for a
varying standard deviation $\sigma$
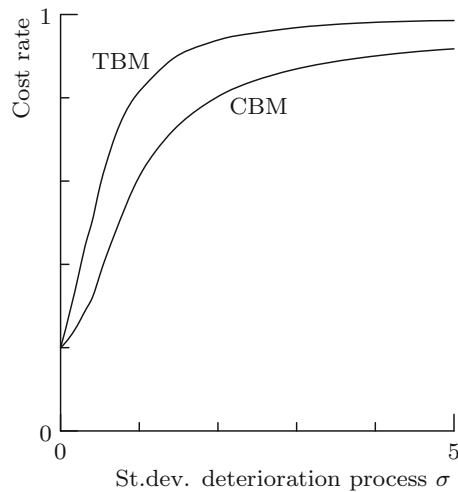of the gamma deterioration
process



Figure 6 shows the effect of the level of volatility $\sigma$ of the gamma deterioration
process on the cost rates of the optimal policies. For very low levels of volatility
there is almost no randomness in the moment of failure, and both CBM and TBM
are very effective, i.e., both are able to carry out preventive maintenance just before
failure. For very high levels of volatility, on the other hand, failure is almost always
caused by a sudden very large deterioration increment. Both the CBM and the
TBM policy cannot prevent this from happening. Note that, in this case, the lifetime
distribution is close to an exponential distribution, and that the optimal age-based
maintenance policy is to never carry out preventive maintenance (because of the
constant failure rate). The benefit of CBM compared to TBM is largest for moderate
levels of volatility.

**Fig. 7** The cost rate under the optimal CBM and the optimal TBM policy for a varying preventive maintenance cost $c_{pm}$
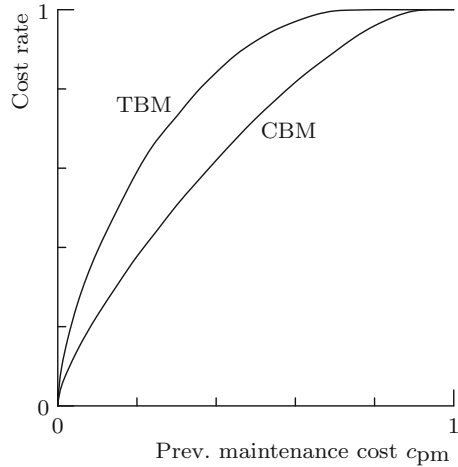
Figure 7 shows the effect of the preventive maintenance cost $c_{pm}$ on the cost rates of the optimal policies. For extremely small preventive maintenance costs, both policies will use a very high maintenance frequency (at very low cost), and almost no failures will happen. This results in a very low cost rate for both policies. For extremely high preventive maintenance costs, carrying out preventive maintenance is not beneficial anymore, and the cost rates of both policies are very high. Again, the cost saving of CBM as opposed to TBM is largest for moderate preventive maintenance costs.

De Jonge et al. (2017) continue to consider the effect of various practical factors that influence the benefit of condition-based maintenance compared to time-based maintenance. The factors that they consider are a required planning time that is needed to carry out preventive maintenance, noise in the observed deterioration information, and uncertainty in the deterioration level at which failure occurs.

In practice there is often a planning time needed between initiating and performing maintenance. Here we assume that a fixed planning time $s$ is required for carrying out preventive maintenance. Furthermore, if failure occurs during the planning time we assume that corrective maintenance will be carried out immediately and that only the high corrective maintenance cost is incurred. The preventive maintenance cost of the maintenance action that was already planned does not need to be paid anymore. We note that a planning time does not influence the time-based maintenance policy. However, for the condition-based maintenance policy, the decision is no longer to determine the deterioration level at which preventive maintenance should be carried out, but it is now the deterioration level at which preventive maintenance should be planned. During the planning time the condition information cannot be used anymore, and, as a consequence, the performance of the condition-based maintenance policy decreases. Figure 8 shows the cost rates of both policies as a function of the planning time $s$. When the planning

**Fig. 8** The cost rate under the optimal CBM and the optimal TBM policy for a varying planning time $s$
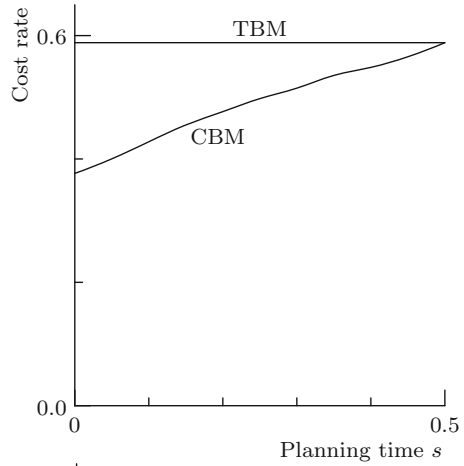


**Fig. 9** The cost rate under the optimal CBM and the optimal TBM policy for a varying level of noise $\sigma_p$ in the condition monitoring
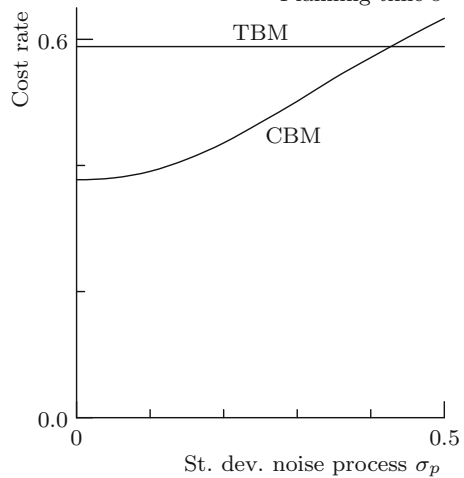


time equals the optimal maintenance age of the time-based maintenance policy, all benefits of condition-based maintenance have vanished.

Another factor that is likely to exist in practice is imperfect condition information due to noise. The difference between the actual deterioration level and the observed deterioration level has been modeled by a Brownian motion, multiplied by a parameter $\sigma_p$. The value of $\sigma_p$ can be interpreted as a measure for the amount of noise. Because the time-based maintenance policy does not use any condition information, noise does not influence the performance of this policy. Condition-based maintenance, on the other, is negatively influenced by imperfect condition monitoring because the obtained information has a lower value. Figure 9 shows the optimal cost rates of both policies as a function of the amount of uncertainty $\sigma_p$. We observe that small amounts of noise only have a minor influence. However, if the amount of noise is substantial, it can even be the case the obtained condition information should not be used at all anymore.

**Fig. 10** The cost rate under the optimal CBM and the optimal TBM policy for a varying level of uncertainty $\sigma_f$ in the deterioration failure level
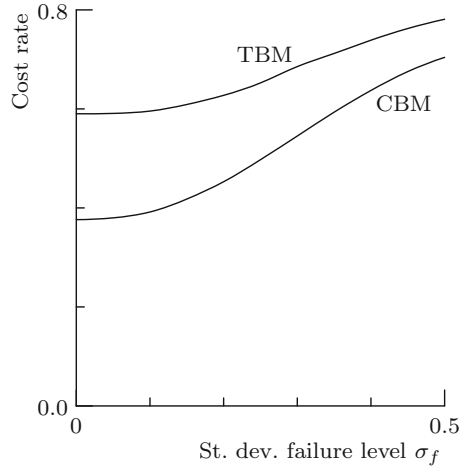
Studies on condition-based maintenance typically assume that failure occurs when a certain fixed level of deterioration is exceeded. In practice, however, there are also many situations where this assumption is not realistic. The randomness in the failure deterioration level has been modeled by a normal distribution with mean 1. The standard deviation $\sigma_f$ of this normal distribution can be seen as a measure for the amount of uncertainty in the failure level. In contrast to the imperfect condition information does the random failure level also affect the time-based maintenance policy. Randomness in the failure level leads to a higher variance in the time until failure, which has a negative impact on the performance of time-based maintenance. The condition-based maintenance policy also suffers from an uncertain failure level as it lowers the value of the condition information. Figure 10 shows that the effect on condition-based maintenance is larger than on time-based maintenance, implying that the benefit of condition-based maintenance is reduced if there is uncertainty in the failure level.

When deciding to switch from time-based maintenance to condition-based maintenance it is important to assess whether the benefits outweigh the additional costs for monitoring equipment and for collecting, storing, and analyzing condition data. It is important that both the volatility of the deterioration process and the cost of preventive maintenance compared to that of corrective maintenance are not extremely low or extremely high. Furthermore, it is important to realize that a required planning time, imperfect condition monitoring, and an uncertain failure level negatively impact the cost saving of condition-based maintenance as opposed to time-based maintenance.

## 3.3 Gamma Deterioration Process with Periodic Inspections

In this section we reconsider the setting of Park (1988), in which a periodic inspection policy is considered for a component that deteriorates according to a stationary increasing continuous-time continuous-state deterioration process. The stationary gamma process is an example of such a process. If an inspection reveals a deterioration level that exceeds a certain threshold level, an immediate preventive maintenance action is carried out. Failure is assumed to occur if a certain fixed failure threshold $L$ is exceeded. Failures are assumed to be self-announcing and are followed by an immediate corrective maintenance action. Furthermore, the inspection schedule is reset after a failure. Both types of maintenance are assumed to make the component as-good-as-new, and to require a negligible amount of time. The cost of preventive maintenance is denoted by $c_{pm}$, the cost of corrective maintenance by $c_{cm}$, and the cost of an inspection by $c_i$. We make the reasonable assumptions that $c_i < c_{pm} < c_{cm}$ and that $c_i + c_{pm} < c_{cm}$.

The maintenance policy in the above setting is described by two decision variables, the time between two consecutive inspections, denoted by $T$, and the preventive maintenance deterioration threshold, denoted by $M$. Initially, we consider the time between inspections as fixed in our analysis, and, for ease of notation, we scale time such that the time between two consecutive inspections is 1. In other words, the $i$th inspection is performed at time $i$. Later on, the time between inspections can be varied to investigate how this influences the optimal cost rate, and to search for the optimal inspection interval.

Given the fixed inspection interval 1, the aim is to obtain an expression for the cost rate $\eta(M)$ as a function of the preventive maintenance threshold $M$. Because the component is as-good-as-new after each maintenance action, standard renewal theory can be applied. We call the time between two consecutive maintenance actions a cycle, and we calculate the cost rate $\eta(M)$ as the mean cost per cycle, denoted by $C(M)$, divided by the mean cycle length, denoted by $D(M)$. That is,

$$\eta(M) = \frac{C(M)}{D(M)}.$$

We will continue to derive expressions for $C(M)$ and $D(M)$, both as a function of the preventive maintenance threshold $M$, which can be evaluated numerically.

We will denote the deterioration process by $X(t)$ with $X(0) = 0$. We let $G_t(x)$ denote the distribution function of the deterioration level at time $t$, i.e., $G_t(x)$ equals the probability that the deterioration level has not exceeded $x$ at time $t$:

$$G_t(x) = P(X(t) < x).$$

We have that $G_0(x) = 1$ for all $x \geq 0$, and $G_t(0) = 0$ for all $t > 0$. The derivative of $G_t(x)$ with respect to $x$ is the density function of the deterioration level at time $t$ and will be denoted by $g_t(x)$:

$$g_t(x) = \frac{\mathrm{d}}{\mathrm{d}x} G_t(x).$$

We use the following expression for the mean cost per cycle $C(M)$:

$$C(M) = c_{\mathrm{pm}} + \sum_{i=1}^{\infty} i \cdot c_{\mathrm{i}} \cdot \mathrm{P}\,(\mathrm{PM\ at\ inspection}\ i)$$

$$+ \sum_{i=1}^{\infty} ((i-1)c_{\mathrm{i}} + c_{\mathrm{cm}} - c_{\mathrm{pm}}) \cdot \mathrm{P}\,(\mathrm{Failure\ between\ inspections}\ i-1\ \mathrm{and}\ i)\,.$$

Thus, we first incur the preventive maintenance cost $c_{\mathrm{pm}}$ and subtract it if a cycle ends with failure. The mean cost per cycle can be written as

$$C(M) = c_{\mathrm{pm}} + \sum_{i=1}^{\infty} i \cdot c_{\mathrm{i}} \cdot \mathrm{P}\,(X(i-1) \le M \text{ and } M < X(i) \le L)$$

$$+ \sum_{i=1}^{\infty} ((i-1)c_{\mathrm{i}} + c_{\mathrm{cm}} - c_{\mathrm{pm}}) \cdot \mathrm{P}\,(X(i-1) \le M \text{ and } X(i) > L)\,.$$

Because the deterioration level at time 0 is degenerate ($X(0) = 0$) we take the first term out of the two summations. Furthermore, by letting $\Delta X_i = X(i) - X(i-1)$ denote the additional amount of deterioration between inspection $i-1$ and inspection $i$, it follows that $C(M)$ can be written as

$$C(M)$$
$$= c_{\mathrm{pm}} + c_{\mathrm{i}} \cdot \mathrm{P}\,(M < X(1) \le L) + (c_{\mathrm{cm}} - c_{\mathrm{pm}}) \cdot \mathrm{P}\,(X(1) > L)$$

$$+ \sum_{i=2}^{\infty} i \cdot c_{\mathrm{i}} \cdot \mathrm{P}\,(X(i-1) \le M \text{ and } M < X(i-1) + \Delta X_i \le L)$$

$$+ \sum_{i=2}^{\infty} ((i-1)c_{\mathrm{i}} + c_{\mathrm{cm}} - c_{\mathrm{pm}}) \cdot \mathrm{P}\,(X(i-1) \le M \text{ and } X(i-1) + \Delta X_i > L)$$

$$= c_{\mathrm{pm}} + c_{\mathrm{i}} \cdot \mathrm{P}\,(M < X(1) \le L) + (c_{\mathrm{cm}} - c_{\mathrm{pm}}) \cdot \mathrm{P}\,(X(1) > L)$$

$$+ \sum_{i=2}^{\infty} i \cdot c_{\mathrm{i}} \cdot \mathrm{P}\,(X(i-1) \le M \text{ and } M - X(i-1) < \Delta X_i \le L - X(i-1))$$

$$+ \sum_{i=2}^{\infty} ((i-1)c_{\mathrm{i}} + c_{\mathrm{cm}} - c_{\mathrm{pm}}) \cdot \mathrm{P}\,(X(i-1) \le M \text{ and } \Delta X_i > L - X(i-1))\,.$$

Because the deterioration level $X(i-1)$ at time $i-1$ is independent of the additional amount of deterioration $\Delta X_i$ between time $i-1$ and time $i$, and because the density

function of $\Delta X_i$ equals $g_1$, we have that $C(M)$ can be written as

$$C(M) = c_{\text{pm}} + c_{\text{i}} \cdot P(M < X(1) \leq L) + (c_{\text{cm}} - c_{\text{pm}}) \cdot P(X(1) > L)$$

$$+ \sum_{i=2}^{\infty} i \cdot c_{\text{i}} \cdot \int_0^M g_{i-1}(x) \int_{M-x}^{L-x} g_1(y) \, \mathrm{d}y \, \mathrm{d}x$$

$$+ \sum_{i=2}^{\infty} ((i-1)c_{\text{i}} + c_{\text{cm}} - c_{\text{pm}}) \cdot \int_0^M g_{i-1}(x) \int_{L-x}^{\infty} g_1(y) \, \mathrm{d}y \, \mathrm{d}x$$

$$= c_{\text{pm}} + c_{\text{i}}(G_1(L) - G_1(M)) + (c_{\text{cm}} - c_{\text{pm}})(1 - G_1(L))$$

$$+ \sum_{i=2}^{\infty} i \cdot c_{\text{i}} \cdot \int_0^M g_{i-1}(x)(G_1(L - x) - G_1(M - x)) \, \mathrm{d}x$$

$$+ \sum_{i=2}^{\infty} ((i-1)c_{\text{i}} + c_{\text{cm}} - c_{\text{pm}}) \cdot \int_0^M g_{i-1}(x)(1 - G_1(L - x)) \, \mathrm{d}x.$$

By rearranging the two sums and combining terms with the variable $i$ in one summation and without it in another summation we obtain

$$C(M) = c_{\text{pm}} + c_{\text{i}}(G_1(L) - G_1(M)) + (c_{\text{cm}} - c_{\text{pm}})(1 - G_1(L))$$

$$+ \sum_{i=2}^{\infty} i \cdot c_{\text{i}} \int_0^M g_{i-1}(x)(1 - G_1(M - x)) \, \mathrm{d}x$$

$$+ \sum_{i=2}^{\infty} (c_{\text{cm}} - c_{\text{pm}} - c_{\text{i}}) \cdot \int_0^M g_{i-1}(x)(1 - G_1(L - x)) \, \mathrm{d}x$$

$$= c_{\text{pm}} + c_{\text{i}}(G_1(L) - G_1(M)) + (c_{\text{cm}} - c_{\text{pm}})(1 - G_1(L))$$

$$+ c_{\text{i}} \sum_{i=2}^{\infty} i \int_0^M g_{i-1}(x) \, \mathrm{d}x - c_{\text{i}} \sum_{i=2}^{\infty} i \int_0^M g_{i-1}(x) G_1(M - x) \, \mathrm{d}x$$

$$+ (c_{\text{cm}} - c_{\text{pm}} - c_{\text{i}}) \sum_{i=2}^{\infty} \int_0^M g_{i-1}(x) \, \mathrm{d}x$$

$$- (c_{\text{cm}} - c_{\text{pm}} - c_{\text{i}}) \sum_{i=2}^{\infty} \int_0^M g_{i-1}(x) G_1(L - x) \, \mathrm{d}x.$$

By realizing that $\int_0^M g_{i-1}(x)G_1(M - x) \, \mathrm{d}x$ equals the probability that the deterioration level at time $i - 1$ is below $M$, and that it is still below $M$ one time period later, we have that

$$\int_0^M g_{i-1}(x)G_1(M-x)\,\mathrm{d}x = \int_0^M g_i(x)\,\mathrm{d}x = G_i(M),$$

which allows us to write $C(M)$ as

$$C(M) = c_{\mathrm{pm}} + c_i(G_1(L) - G_1(M)) + (c_{\mathrm{cm}} - c_{\mathrm{pm}})(1 - G_1(L))$$

$$+ c_i \sum_{i=2}^{\infty} i G_{i-1}(M) - c_i \sum_{i=2}^{\infty} i G_i(M)$$

$$+ (c_{\mathrm{cm}} - c_{\mathrm{pm}} - c_i) \sum_{i=2}^{\infty} G_{i-1}(M) - (c_{\mathrm{cm}} - c_{\mathrm{pm}} - c_i)$$

$$\times \int_0^M \sum_{i=2}^{\infty} g_{i-1}(x)G_1(L-x)\,\mathrm{d}x$$

$$= c_{\mathrm{pm}} + c_i(G_1(L) - G_1(M)) + (c_{\mathrm{cm}} - c_{\mathrm{pm}})(1 - G_1(L))$$

$$+ c_i G_1(M) + c_i \sum_{i=1}^{\infty} G_i(M)$$

$$+ (c_{\mathrm{cm}} - c_{\mathrm{pm}} - c_i) \sum_{i=1}^{\infty} G_i(M) - (c_{\mathrm{cm}} - c_{\mathrm{pm}} - c_i)$$

$$\times \int_0^M \sum_{i=1}^{\infty} g_i(x)G_1(L-x)\,\mathrm{d}x.$$

For ease of notation we let the function $H(x)$ be defined as

$$H(x) = \sum_{i=1}^{\infty} G_i(x),$$

and the function $h(x)$ as the derivative of $H(x)$, i.e.,

$$h(x) = \frac{\mathrm{d}}{\mathrm{d}x} H(x) = \sum_{i=1}^{\infty} g_i(x).$$

We then have that $C(M)$ can be expressed as

$$C(M) = c_{\mathrm{pm}} + c_i(G_1(L) - G_1(M)) + (c_{\mathrm{cm}} - c_{\mathrm{pm}})(1 - G_1(L))$$
$$+ c_i G_1(M) + c_i H(M)$$
$$+ (c_{\mathrm{cm}} - c_{\mathrm{pm}} - c_i)H(M) - (c_{\mathrm{cm}} - c_{\mathrm{pm}} - c_i) \int_0^M h(x)G_1(L-x)\,\mathrm{d}x,$$

which can be rewritten to our following final expression for the mean cost per cycle $C(M)$:

$$C(M) = c_{\text{cm}} + (c_{\text{cm}} - c_{\text{pm}})H(M) - (c_{\text{cm}} - c_{\text{pm}} - c_{\text{i}})(G_1(L)$$

$$+ \int_0^M h(x)G_1(L - x)\,dx).$$

We will now continue with the mean cycle length $D(M)$, which can be expressed as

$$D(M) = \sum_{i=1}^{\infty} i \cdot P\,(\text{PM at inspection } i)$$

$$+ \sum_{i=0}^{\infty} E\,(\text{Time until failure between inspections } i \text{ and } i + 1)\,.$$

Similar to the determination of $C(M)$, we can show that

$$\sum_{i=1}^{\infty} i \cdot P\,(\text{PM at inspection } i)$$

$$= G_1(L) - G_1(M) + \sum_{i=2}^{\infty} i \cdot \int_0^M g_{i-1}(x)(G_1(L - x) - G_1(M - x))\,dx$$

$$= G_1(L) - G_1(M) + \sum_{i=1}^{\infty}(i + 1) \int_0^M g_i(x)(G_1(L - x) - G_1(M - x))\,dx. \qquad (2)$$

We let $F_x(t)$ and $f_x(t)$ respectively denote the distribution and density function of the time $t$ at which deterioration level $x$ is reached. We have

$$F_x(t) = P(X(t) \geq x) = 1 - P(X(t) \leq x) = 1 - G_t(x).$$

We can now write

$$\sum_{i=0}^{\infty} E\,(\text{Time until failure between inspections } i \text{ and } i + 1)$$

$$= \int_0^1 t f_L(t)\,dt + \sum_{i=1}^{\infty} \int_0^M \int_0^1 (i + t)g_i(x) f_{L-x}(t)\,dt\,dx$$

$$= F_L(1) - \int_0^1 F_L(t)\,dt + \sum_{i=1}^{\infty} \int_0^M \int_0^1 i g_i(x) f_{L-x}(t)\,dt\,dx$$

$$+ \sum_{i=1}^{\infty} \int_0^M \int_0^1 t g_i(x) f_{L-x}(t) \, \mathrm{d}t \, \mathrm{d}x$$

$$= \int_0^1 G_t(L) \, \mathrm{d}t - G_1(L) + \sum_{i=1}^{\infty} \int_0^M i g_i(x) F_{L-x}(1) \, \mathrm{d}x$$

$$+ \int_0^M \int_0^1 t h(x) f_{L-x}(t) \, \mathrm{d}t \, \mathrm{d}x$$

$$= \int_0^1 G_t(L) \, \mathrm{d}t - G_1(L) + \sum_{i=1}^{\infty} i \int_0^M g_i(x)(1 - G_1(L-x)) \, \mathrm{d}x$$

$$+ \int_0^M h(x) \left( \int_0^1 G_t(L-x) \, \mathrm{d}t - G_1(L-x) \right) \, \mathrm{d}x. \tag{3}$$

By combining (2) and (3) it can be shown that the mean cycle length $D(M)$ equals

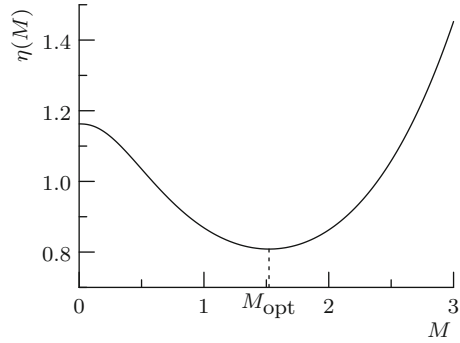$$D(M) = \int_0^1 G_t(L) \, \mathrm{d}t + \int_0^M h(x) \int_0^1 G_t(L-x) \, \mathrm{d}t \, \mathrm{d}x.$$

The cost rate $\eta(M)$ as a function of the preventive maintenance threshold is thus equal to

$$\eta(M) = \frac{C(M)}{D(M)}$$

$$= \frac{c\mathrm{cm} + (c\mathrm{cm} - c\mathrm{pm})H(M) - (c\mathrm{cm} - c\mathrm{pm} - c_\mathrm{i})(G_1(L) + \int_0^M h(x)G_1(L-x) \, \mathrm{d}x)}{\int_0^1 G_t(L) \, \mathrm{d}t + \int_0^M h(x) \int_0^1 G_t(L-x) \, \mathrm{d}t \, \mathrm{d}x}.$$

$$\tag{4}$$

We will now consider the specific stationary gamma deterioration process with parameter values $a = 2.5$ and $b = 0.5$. Furthermore, we assume a breakdown deterioration level $L = 4$, a corrective maintenance cost $c\mathrm{cm} = 10$, a preventive maintenance cost $c\mathrm{pm} = 1$, and an inspection cost $c_\mathrm{i} = 0.1$. Note that the inspection interval is still fixed at 1. Figure 11 shows the cost rate $\eta(M)$ as function of the preventive maintenance threshold $M$. It turns out to be optimal to carry out preventive maintenance if an inspection reveals a deterioration level of at least $M_\mathrm{opt} = 1.52$. The corresponding cost rate equals $\eta(M_\mathrm{opt}) = 0.81$.

In the case of a stationary gamma deterioration process, (4) basically provides us with a formula $\eta(M, L, a, b, c\mathrm{cm}, c\mathrm{pm}, c_\mathrm{i})$ for the cost rate, in which $L, a, b, c\mathrm{cm}$, $c\mathrm{pm}$, and $c_\mathrm{i}$ are model parameters. For a stationary gamma deterioration process with parameters $a$ and $b$, the deterioration increment during a time period of length $T$ is gamma distributed with parameters $aT$ and $b$. Therefore, for an inspection interval with an arbitrary length $T$, the cost rate $\bar{\eta}$ can easily be expressed in terms

**Fig. 11** The cost rate $\eta(M)$ as a function of the preventive maintenance threshold $M$



**Fig. 12** The cost rate $\bar{\eta}(T, M)$ for various inspection intervals $T$ and as a function of the preventive maintenance threshold $M$
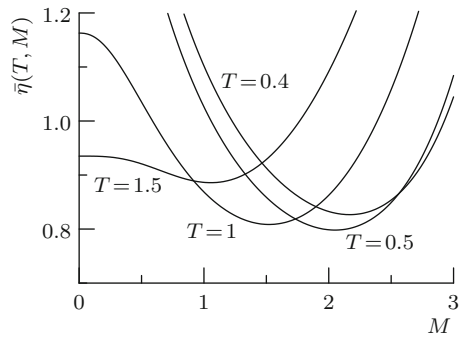


of the cost rate $\eta$ for an inspection interval with length 1:

$$\bar{\eta}(T, M, L, a, b, c_{cm}, c_{pm}, c_i) = \frac{\eta(M, L, aT_i, b, c_{cm}, c_{pm}, c_i)}{T}.$$

Figure 12 shows the cost rate $\eta(M, T)$ for various inspection intervals $T$, again as a function of the preventive maintenance threshold $M$. Based on this figure it can be concluded that the optimal inspection interval should be somewhere between 0.4 and 1. If we optimize $\eta(M, T)$ numerically over both $M$ and $T$, we find that the optimal inspection interval equals $T_{opt} = 0.68$, and that preventive maintenance should be carried out if an inspection reveals a deterioration level of at least $M_{opt} = 1.85$. Thus, by allowing an inspection interval with length different from 1, it is optimal to inspect the component more frequently, and, as a consequence, the preventive maintenance threshold will increase. The corresponding optimal cost rate decreases to $\bar{\eta}(T_{opt}, M_{opt}) = 0.78$.

## 3.4 Discretizing Continuous-Time Continuous-State Deterioration Processes

The drawback of modeling deterioration by a continuous-time continuous-state stochastic process is its complicated analytical tractability. The maintenance policies that we have considered in Sect. 3.2 are for instance difficult to evaluate numerically. The main reason for this is the overshoot behavior of the gamma process that is caused by the fact that it makes jumps. The analysis in Sect. 3.2 is therefore based on simulation.

An alternative method that can be used to analyze maintenance policies for a component that deteriorates according to a continuous-time continuous-state process is by discretizing this process. De Jonge (2019) presents an approach for discretizing stationary non-decreasing continuous-time continuous-state deterioration processes into discrete-time Markov chains with stationary increments. The first step of this approach is to discretize the continuous time into discrete time steps with a certain length $\Delta t$. Furthermore, the deterioration levels between 0 and the failure level $L$ are subdivided into $m$ deterioration intervals $x_k$, $k = 1, \ldots, m$. These intervals correspond to states $1, \ldots, m$ in the Markov chain. All deterioration levels above $L$ are combined into the failed state $m + 1$. The transition probabilities of the Markov chain are calculated based on the assumption that the deterioration level is uniformly distributed on a certain interval $x_k$ when it is within this interval at an arbitrary moment in time.

As an example, if we consider a stationary gamma deterioration process with parameters $a = 2$ and $b = 0.2$, a failure threshold level $L = 1$, a number of deterioration states before failure of $m = 4$, and time steps with length $\Delta t = 0.1$, we obtain the following transition probability matrix for the discrete-time Markov chain:

$$P = \begin{pmatrix} 0.861847 & 0.120864 & 0.013780 & 0.002713 & 0.000795 \\ 0.000000 & 0.861847 & 0.120864 & 0.013780 & 0.003509 \\ 0.000000 & 0.000000 & 0.861847 & 0.120864 & 0.017288 \\ 0.000000 & 0.000000 & 0.000000 & 0.861847 & 0.138153 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 1.000000 \end{pmatrix}.$$

De Jonge (2019) also points out how the initial maintenance policy considered in Sect. 3.2 can be evaluated based on the discretized deterioration process and on matrix algebra. Because failed components will remain failed as long as no maintenance is carried out, the Markov chain with transition probability matrix $P$ is an absorbing Markov chain with state $m + 1$ the absorbing state. The matrix $P$ can be written as

$$P = \begin{pmatrix} Q & \mathbf{r} \\ \mathbf{0} & 1 \end{pmatrix},$$

in which $Q$ is an $m \times m$ matrix. The probability of going from a deterioration state $i \leq m$ to a deterioration state $j \leq m$ in exactly $k$ time steps is equal to entry $(i, j)$ of the matrix $Q^k$. The fundamental matrix $R$ is given by

$$R = \sum_{k=0}^{\infty} Q^k = (I_m - Q)^{-1},$$

in which entry $(i, j)$ equals the expected number of time periods that the process is in state $j$ before it is being absorbed, given that it started in state $i$. After carrying out maintenance the component is as-good-as-new and the expected time until failure equals $\sum_j R_{1j}$.

We let $M \in \{1, \ldots, m\}$ denote the preventive maintenance threshold, and $\eta(M)$ the corresponding cost rate. Standard renewal theory can again be used to calculate this cost rate. We let $C(M)$ denote the mean cost per maintenance action and $D(M)$ the mean time until maintenance. Thus,

$$\eta(M) = \frac{C(M)}{D(M)}.$$

Because the deterioration process is non-decreasing, we have that $R_{1j}$, $j < M$, is also the expected number of time periods that the deterioration level is $j$ before reaching a deterioration level of at least $M$, i.e., before maintenance is carried out. This results in the following expression for the mean time until maintenance:

$$D(M) = \sum_{j < M} R_{1j}.$$

Furthermore, because the probability of failure is $P_{j,m+1}$ if the deterioration level is $j$, it follows that the probability that a cycle ends with failure equals $\sum_{j < M} R_{1j} P_{j,m+1}$, implying that the mean cost per maintenance action equals

$$C(M) = c_{\mathrm{pm}} + (c_{\mathrm{cm}} - c_{\mathrm{pm}}) \sum_{j < M} R_{1j} P_{j,m+1}.$$

Based on the above we now have the following expression for cost rate:

$$\eta(M) = \frac{C(M)}{D(M)} = \frac{c_{\mathrm{pm}} + (c_{\mathrm{cm}} - c_{\mathrm{pm}}) \sum_{j < M} R_{1j} P_{j,m+1}}{\sum_{j < M} R_{1j}}.$$

By choosing a sufficiently high number of deterioration states $m$ in the discretization, this formula provides us with a smooth graph of the cost rate as a function of the preventive maintenance threshold $M$. Figure 13 shows this cost rate for the case that we have considered in Sect. 3.2, i.e., a stationary gamma deterioration process with parameters $a = 5$ and $b = 0.22$, a failure threshold $L = 1$, and cost parameters

**Fig. 13** The cost rate $\eta(M)$ as a function of the preventive maintenance threshold $M$
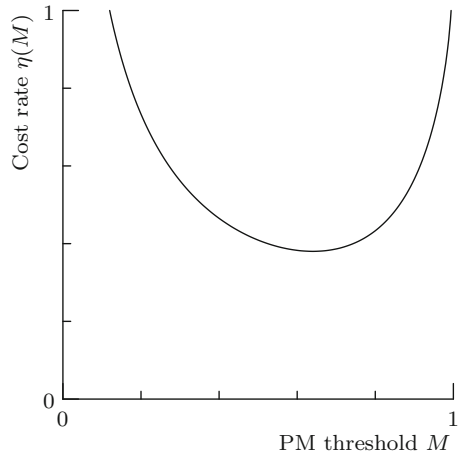
$c_{pm} = 0.2$ and $c_{cm} = 1$. Furthermore, $m = 100$ deterioration states before failure have been considered. We observe that this graph is virtually identical to the graph of CBM in Fig. 5. The main advantage of this approach is that we avoid the long calculation times that are required for simulation.

## 3.5 Aperiodic Inspections

Modeling deterioration by a discrete-time Markov chain is also useful when aiming to determine optimal policies by using the framework of Markov decision processes. This methodology is for instance applicable for determining maintenance policies with aperiodic inspections. We again consider a single component that deteriorates according to a discrete-time Markov chain with transition probability matrix $P$. There are $m$ deterioration states before failure and a state $m + 1$ that represents failure. Failures are assumed to be self-announcing; all other deterioration states can only be observed by an inspection. Inspections can be performed at the start of each time period, the cost of an inspection is denoted by $c_i$, and an inspection is assumed to take a negligible amount of time. Furthermore, also at the start of each time period, preventive maintenance can be performed. This can done immediately after an inspection, based on the observed deterioration level, or without performing an inspection first. When failure occurs, corrective maintenance should be carried out immediately. Both preventive and corrective maintenance are assumed to take a negligible amount of time and to bring the component back to the as-good-as-new state. The costs of a preventive and of a corrective maintenance action are denoted by $c_{pm}$ and $c_{cm}$, respectively. This setting is also considered by Maillart (2006), in particular for a small number of deterioration states.

The optimal inspection and maintenance decisions can be determined by formulating the above as a Markov decision process. Because the exact deterioration state of the components is uncertain as long as no inspection or maintenance is carried out, and because this uncertainty cannot be ignored, it is appropriate to formulate the problem as a partially observable Markov decision process (Monahan, 1982). A partially observable Markov decision process is a generalization of the standard Markov decision process, and can be formulated as a Markov decision process with an enlarged state space, namely the space of probability distributions over the underlying states. The states of a partially observable Markov decision process are typically called either knowledge states or belief states.

In the setting that we consider it is convenient to denote the knowledge states by, for instance, $\theta^{i,j}$, in which $i$ denotes the last observed deterioration level, and $j$ denotes the number of time periods ago that this deterioration level has been observed, $i = 1, \ldots, m$, $j = 0, 1, \ldots$. Thus, $\theta^{i,0}$ denotes the knowledge state if it is known with certainty that the current deterioration level is $i$. For $j > 0$ the actual deterioration level is uncertain, and in general, the probability of a sudden failure increases both in $i$ and in $j$. The exact probabilities can be calculated based on the transition probability matrix $P$ of the Markov chain. Another remark is that the number of knowledge state is infinite. However, we can fairly choose a sufficiently large $N$ for which we can be reasonably sure that, under the optimal policy, the time between two consecutive actions (either inspection or maintenance) will never exceed $N$ periods. This results in a finite number of states $\theta^{i,j}$, $i = 1, \ldots, m$, $j = 0, 1, \ldots, N$.

In any state the optimal action will always be either to do nothing, to carry out an inspection, or to perform preventive maintenance. Corrective maintenance is performed immediately when failure occurs and is therefore not really considered as an action. In other words, if failure occurs, we incur cost $c_{cm}$ and we immediately move to state $\theta^{1,0}$. If we are in state $\theta^{1,0}$ the component is as-good-as-new with certainty, the optimal action will thus be to do nothing. For states $\theta^{i,0}$, $i = 2, \ldots, m$, the deterioration level is also known with certainty, implying that the optimal action will be either to do nothing or to carry out preventive maintenance. In all other states, any of the three actions can be chosen. Based on this reasoning, the value iteration algorithm (Puterman, 1994) can be applied, and the optimal inspection and maintenance policy can be determined.

We will continue to consider an example. We consider a component that deteriorates according to a stationary gamma deterioration process with parameters $a = 0.5$ and $b = 0.25$, and with failure deterioration level $L = 1$. We will discretize this gamma process by using the approach in Sect. 3.4, and we will use $m = 50$ deterioration states before the failed state, and time steps with length $\Delta t = 0.1$. The cost of corrective maintenance is $c_{cm} = 5$, that of preventive maintenance $c_{pm} = 1$, and that of an inspection $c_i = 0.1$.

Figure 14 shows the optimal inspection and maintenance policy. The horizontal axis shows the last revealed deterioration state (for the discrete-state deterioration process), and the vertical axis the number of periods between observing this state and the next preventive maintenance action or inspection. If an inspection reveals a
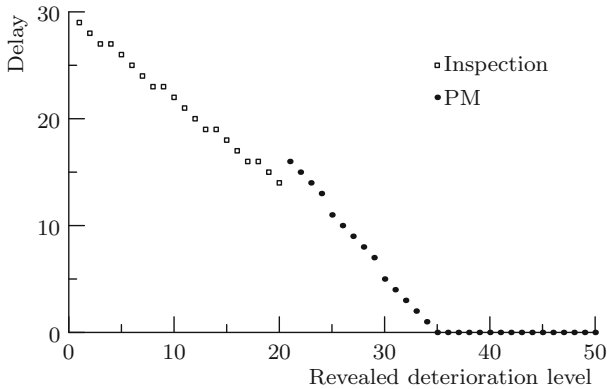
**Fig. 14** Optimal action and delay time as a function of the currently revealed deterioration level

deterioration state of at most 20, we observe that a new inspection will be scheduled. The time until this next inspection is decreasing in the observed deterioration state, resulting in a dynamic aperiodic maintenance policy. If a deterioration state of at least 21 is revealed by an inspection, preventive maintenance will be carried out, either immediately or after a certain number of time periods. For deterioration states 21–34 an immediate failure is not that likely, but scheduling another inspection is not cost effective. In these cases, a delayed preventive maintenance action will be scheduled, with a delay time that is decreasing in the observed deterioration state. For an observed deterioration state of at least 35, the risk of a failure is deemed too high, and an immediate preventive maintenance action will be carried out.

## 4   Concluding Remarks

We have considered maintenance policies for non-repairable components, i.e., maintenance interventions can be seen as a replacement of the component. We started to consider two time-based maintenance policies, viz., age-based maintenance and block-based maintenance. For the age-based maintenance policy we have considered the effect of uncertainty in the scale parameter of the lifetime distribution on the optimal preventive maintenance age. This setting could be extended to uncertainty in other parameters of the lifetime distribution as well, or uncertainty in the parametric distribution itself (model uncertainty). For the block-based maintenance policy we have mainly focused on the optimal maintenance interval under random usage of the component. Suitable extensions of this setting would be to consider multiple component speeds, instead of only on and off, and some flexibility in the usage of the component. Furthermore, the effect of uncertainty in the lifetime distribution is also of interest in settings with block-based maintenance.

We continued to consider condition-based maintenance polices. First, we mentioned some recent developments in delay-time modeling. After that, we have adopted a continuously monitored stationary gamma deterioration process and we have considered the performance of condition-based maintenance as opposed to time-based maintenance. This analysis was based on simulation and studied the effect of the volatility of the deterioration process and of the relative cost of preventive maintenance. Furthermore, the presence of a planning time, of noise in the obtained deterioration information, and of uncertainty in the lifetime distribution have been considered. After this, we have considered a stationary gamma deterioration process combined with periodic inspections. We have obtained mathematical expressions to simultaneously optimize the inspection interval and the preventive maintenance deterioration threshold. Finally, we have provided an approach that can be used to discretize continuous-time continuous-state deterioration processes. We have first used the obtained Markov chain to reconsider the condition-based maintenance policy for a continuously monitored stationary gamma process. This analysis is based on matrix algebra. Thereafter, we have pointed out how the Markov chain and the concept of Markov decision processes can be used to determine optimal aperiodic inspection and maintenance policies.

The models with condition-based maintenance could be extended by considering various types of uncertainty. The parameters of the gamma deterioration process, or even the functional form of the deterioration process could be unknown. Furthermore, the degree of imprecision of the deterioration increments could be uncertain, or the distribution of a random failure level could be unknown. As a final suggestion, random usage of a component or production decisions could also be considered in settings with condition-based maintenance.

# References

Abdel-Hameed M (1975) A gamma wear process. IEEE Trans Reliab R-24(2):152–153

Barlow RE, Proschan F (1965) Mathematical theory of reliability. Wiley, New York

Christer AH (1976) Innovative decision making. In: Bowen KC, White DJ (eds) Proceedings of the NATO Conference on the role and Effectiveness of Theories of Decision in Practice. Hodder and Stoughton, London, pp 368–377

De Jonge B (2019) Discretizing continuous-time continuous-state deterioration processes, with an application to condition-based maintenance optimization. Reliab Eng Syst Saf 188:1–5

De Jonge B, Jakobsons E (2018) Optimizing block-based maintenance under random machine usage. Eur J Oper Res 265(2):703–709

De Jonge B, Dijkstra AS, Romeijnders W (2015) Cost benefits of postponing time-based maintenance under lifetime distribution uncertainty. Reliab Eng Syst Saf 140:15–21

De Jonge B, Klingenberg W, Teunter RH, Tinga T (2015) Optimum maintenance strategy under uncertainty in the lifetime distribution. Reliab Eng Syst Saf 133:59–67

De Jonge B, Teunter RH, Tinga T (2017) The influence of practical factors on the benefits of condition-based maintenance over time-based maintenance. Reliab Eng Syst Saf 158:21–30

Gertsbakh I (2000) Reliability theory, with applications to preventive maintenance. Springer, Berlin

Maillart LM (2006) Maintenance policies for systems with condition monitoring and obvious failures. IIE Trans 38(6):463–475

Monahan GE (1982) A survey of partially observable markov decision processes: theory, models, and algorithms. Manag Sci 28(1):1–16

Park KS (1988) Optimal continuous-wear limit replacement under periodic inspections. IEEE Trans Reliab 37(1):97–102

Puterman ML (1994) Markov decision processes. Wiley, New York

Tokumoto S, Dohi T, Yun WY (2014) Bootstrap confidence interval of optimal age replacement policy. Int J Reliab Qual Saf Eng 21(4):1450018

Van Oosterom CD, Elwany AH, Çelebi D, Van Houtum G-J (2014) Optimal policies for a delay time model with postponed replacement. Eur J Oper Res 232(1):186–197

Wang H, Wang W, Peng R (2017) A two-phase inspection model for a single component system with three-stage degradation. Reliab Eng Syst Saf 158(1):31–40