Probabilistic models for life-cycle performance of deteriorating structures: review and future directions

Dan M Frangopol¹, Maarten-Jan Kallen² and Jan M van Noortwijk²
¹University of Colorado at Boulder, USA
²HKV Consultants and Delft University of Technology, The Netherlands

Summary
In comparison with the well-researched field of analysis and design of structural systems, the life-cycle performance prediction of these systems under no maintenance as well as under various maintenance scenarios is far more complex, and is a rapidly emergent field in structural engineering. As structures become older and maintenance costs become higher, different agencies and administrations in charge of civil infrastructure systems are facing challenges related to the implementation of structure maintenance and management systems based on life-cycle cost considerations. This article reviews the research to date related to probabilistic models for maintaining and optimizing the life-cycle performance of deteriorating structures and formulates future directions in this field.

Key words: maintenance; inspection; life-cycle costing; decision model; optimization; reliability; Markov decision process; renewal model; gamma process

Introduction
The management of individual or groups of structures, requires a systematic approach such that the reliability and condition of the structure(s) can be maintained under budget and/or resource constraints. This means that maintenance and inspection activities should be optimally planned in order to ensure the safe and economic operation of structures. An important concept in maintenance modelling is that of life-cycle costing, where the effects and costs of a particular maintenance policy are considered over the total expected lifetime of structures.

Every maintenance model will somehow try to predict or extrapolate the future performance of the structure in question, whether it be in a deterministic or probabilistic fashion. The current and future states of a structure are associated with various degrees of uncertainty and, therefore, a probabilistic approach is necessary. Many different models have been developed in various fields of application, e.g. industrial, civil, electrical and mechanical engineering. During the past four decades, a large number of papers on maintenance optimization models, mainly focusing on the mathematical aspects, have been published. For a chronological (but unfortunately inherently incomplete) overview, see[1–11]. In this paper a number of probabilistic models for maintaining and optimizing the life-cycle performance of deteriorating structures, with a focus on applications to civil structures and emphasizing highway bridges, are reviewed and future directions are presented.

Each model can roughly be divided into two parts: a deterioration model and a decision model. The deterioration model is used to approximate and predict the actual process of ageing in condition or in reliability. The decision model uses the deterioration model to determine the optimal times of inspection and maintenance. In most cases, a cost-optimal maintenance policy under performance constraints is determined. Maintenance policies can be either periodic (uniform) or they can be aperiodic (nonuniform). Periodic policies are cyclic policies in which the times between maintenance actions are equal.
In this article, a distinction is made between random-variable and stochastic process models. The random-variable models assume that one or more of the variables in the deterioration model is/are random, i.e. a probability distribution is assigned to the uncertain variables. The stochastic process models assume that the deterioration over time is represented by a collection of random variables. In essence, the uncertain deterioration is propagated forward in time. Since the random-variable models consider the uncertain condition at an evaluation time, these models are more static compared with the stochastic process models.

2 Random-variable models

Three random variable models are discussed. First, the failure rate models in which the only random variable is the lifetime itself. Second, the classical reliability index model, where the lifetime distribution follows from a limit state which is a function of one or more (physical) random variables. Third, the condition index model where the lifetime distribution follows, possibly from results of visual inspections. Finally, the time-dependent reliability and condition index models are presented, where a reliability and condition profiles are generated from random variables.

2.1 Failure rate

Deterioration model. A lifetime distribution represents the uncertainty in the time to failure of a component or structure. Let the lifetime \( t \) have a cumulative probability distribution \( F(t) \) with probability density function \( f(t) \), then the failure rate function is defined as [1, Chapter 2]

\[
r(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{F(t)}, \quad t > 0.
\]

A probabilistic interpretation of the failure rate function is that \( r(t) \) represents the probability that a component of age \( t \) will fail in the time interval \([t, t + dt]\). An alternative name for the failure rate is the hazard rate. For deteriorating components or structures, the failure rate is increasing. Lifetime distributions and failure rate functions are especially useful in mechanical and electrical engineering. In these fields, one often considers equipment which can assume at most two states: the functioning state and the failed state. For example, a motor or switch is either working or not. A structure, on the other hand, can be in a range of states depending on its degrading condition. A serious disadvantage of failure rates is that they cannot be observed or measured[12].

Decision model. The area of optimizing maintenance through mathematical models based on lifetime distributions was founded in the early 1960s[1,2]. Well-known decision models of this period are the age replacement model (replacement upon failure or upon reaching a predetermined age \( k \), whichever occurs first) and the block replacement model (replacement upon failure and periodically at the times \( k, 2k, 3k \ldots \) [1, Chapters 3–4]. The age replacement model is one of the most frequently applied maintenance optimization models. Reference can be made to Dekker[8] and Dekker & Scarf[10] for a review of applications.

2.2 Reliability index

Deterioration model. A state function \( g \) is commonly defined as the difference between the structure or component’s resistance \( R \) and the applied load or stress \( S \) : \( g = R - S \). The cross-section, component or structure is considered to operate safely as long as \( g \geq 0 \) and unsafe when \( g < 0 \). The point where \( g = 0 \) is called the limit state and this is where failure is assumed to occur.

In the general case the limit state function \( g(X_1, \ldots, X_n) \) is considered, where the \( X_i \) are \( n \) random variables. The state function will, in many cases, also be a function of time, e.g.,

\[
g = g(t) = g(t, X_1, \ldots, X_n), \quad \text{or of the number of load cycles.}
\]

Note that this model can also be considered as a stochastic process, where the definition of a stochastic process (see Section 3) is strictly applied.

The probability of failure \( \Pr\{g(X) < 0\} \) is the total mass of the joint density for \( g \) which is in the failure region \( \Omega = \{x|g(x) < 0\} \). This mass is given by the \( n \)-dimensional integral:

\[
\Pr\{g(X) < 0\} = \int_{x \in \Omega} f_X(x_1, \ldots, x_n) \, dx_1 \ldots dx_n,
\]

where \( f_X(x) \) is the joint density for \( X = (X_1, \ldots, X_n) \). This integral can usually not be solved analytically and there are a number of approaches available to approximate the probability of failure. Three of the most common approaches are briefly discussed herein. Two of these, the mean value first-order second-moment (MVFOSM) method and the first-order reliability method (FORM), are based on the concept of the reliability index, which is usually denoted by \( \beta \). When the value of \( \beta \) is known, the probability of failure is approximated by calculating the value of the standard normal distribution at \( -\beta \):

\[
\Pr\{g < 0\} = \Phi(-\beta).
\]

The MVFOSM methodology is based on the simple model in which the resistance and the stress in the state function \( g = R - S \) are normally distributed. The output is then also normally distributed with mean \( \mu_g \) and standard deviation \( \sigma_g \) and Cornell[13] named the ratio \( \beta = \mu_g / \sigma_g \) the safety index. This approach has mostly been replaced by FORM due to its restrictive assumptions of normality and linearity. Also, the result does depend on the problem formulation[14].

In a FORM analysis, the input variables are first transformed to standard normal variables (with mean...
zero and standard deviation one) and then $\beta$ (also referred to as the Hasofer–Lind[15] reliability index) is found by determining the distance between the origin of the variable space and the so-called design point. This point represents the worst combination of the model variables which most likely results in failure. Determining the design point is the most tedious part of this methodology. It requires an iterative search algorithm such as the Rackwitz–Fiessler algorithm[16] or it can be found by solving a non-linear quadratic optimization problem[17,18].

The solution to eq. (2) can also be approximated by Monte Carlo simulation (also called Monte Carlo integration). Several alternative simulation-based approaches have been suggested to improve the poor efficiency of this method. Of these alternatives, directional sampling has received most interest in recent years[18–20].

A detailed discussion on the reliability index is beyond the scope of this article. For more on this subject, the reader is referred to the classic textbooks by Thoft-Cristensen and Baker[21], Madsen et al.[22] and Melchers[20].

**Decision model.** Here we focus on how the reliability index is used to determine optimal maintenance policies. Once a suitable solution method for the failure probability given by eq. (2) is selected, a decision model has to be formulated. As suggested by Stewart[23], there are two most suitable approaches to decision analysis: life-cycle costs and risk ranking.

Risk ranking is only useful for the purpose of inspection prioritization at the time of evaluation or at a fixed time in the future. It does not account for the full life-cycle of a structure or component, but only considers its immediate risk. Risk is usually defined as the product of the failure probability with the consequence of failure. Hence, the highest risk items will be those with both a high failure probability and severe consequences. In the case of bridge management, risk ranking could be used to identify the most critical bridges in a network. In the process industry, plant managers use risk ranking to select those components which constitute the highest risk out of thousands in a typical plant. A commonly used tool is the so-called risk matrix, which categorizes all components according to their failure probabilities and their expected failure consequences. A large industry sponsor group has recently developed a full methodology for the evaluation of steel pressure vessels[24]. This methodology includes a simple application of Bayesian updating to assess which type of inspection will be most effective at reducing the uncertainty in the deterioration.

The life-cycle approach is the preferred concept when decision makers are not only concerned with safety, but also with costs. A decision tree is often used for the optimal life-cycle cost planning of inspections and subsequent maintenance actions. If we inspect a component or structure at times $t_0 \leq t_1 \leq t_2 \leq \ldots$ and we decide after each inspection to perform a maintenance action or not, then all possible options can be visualized by a decision tree[25] such as the one in Fig. 1.

After each inspection, the decision is made to perform a maintenance action $A$ (branch labeled with 1) or no action (branch labeled 0). The probability of performing action $A$ will be determined by the state of the component or system. For example: the probability of replacement due to failure is given by the solution to eq. (2). This probability can also be represented by the reliability index $\beta$. One can also define formulations for preventive replacements or other types of maintenance actions. If we are interested in an optimal periodic inspection policy (i.e. the times between two inspections are always equal), then we can calculate the costs associated with the decision tree for each inspection interval and choose the interval which minimizes the expected costs. Although this is a simple solution approach, it does not allow the decision maker to insert constraints on the solution. For this purpose, the optimization problem can be formulated as a nonlinear constrained optimization. Formulations for a single component and for a system of components have been given[25,26]. With this approach a minimum required reliability in the constraints can be easily defined, such that the optimal random or periodic solution ensures a minimum level of safety. This decision tree approach, described in detail[26], has been applied to the inspection of metal fatigue in steel bridges[27], crack measurement and repair in steel pipelines[28], corrosion of steel reinforcement in concrete bridges[26] and periodic inspection planning of a steel-girder bridge[29]. The latter article is the only one which does...
not include a probability of detection of deterioration. The other articles all include this and assume that damage above a certain threshold is perfectly observed. An elegant way of incorporating the probability of detection and inspection uncertainty is presented in Pandey[28], where a Bayesian updating scheme is used. For this purpose, a number of suitable detection functions for cracking in steel bridges are given[30]. The effect of modelling inspection uncertainty is demonstrated by Mori & Ellingwood[31], where the deterioration uncertainty is also updated after every inspection or repair. Both[31] and its second part[32], are good examples of the application of the decision tree optimization approach. A non-Bayesian approach is used in Frangopol et al.[26], which requires about eleven relationships including seven constants and two uncertain variables to incorporate the probability of detection.

### 2.3 Time-dependent reliability index

Owing to the intensive use of the reliability index in code calibration and in reliability-based analysis and design, a time-dependent reliability index approach to maintaining the safety and optimizing the life-cycle performance of deteriorating structures is desirable. Thoft-Christensen[33,34], Nowak et al.[35], Frangopol et al.[36,37], and Kong & Frangopol[38,39], among others, used the concept of the reliability index profile for deteriorating structures without maintenance. The reliability profile is defined as the variation of the reliability index with time. Furthermore, time-dependent reliability index models were developed and applied to extend the service life of deteriorating structures under various maintenance scenarios.

**Deterioration model without maintenance.** Up to now, the bi-linear reliability index profile

$$\beta(t) = \begin{cases} 
\beta_0 & \text{for } 0 \leq t \leq t_1, \\
\beta_0 - \alpha_1(t - t_1) & \text{for } t > t_1 
\end{cases}$$

is applied extensively[33-39]. Recently, nonlinear reliability index deterioration models were proposed by Petcherdchoo et al.[40] as follows:

$$\beta(t) = \begin{cases} 
\beta_0 & \text{for } 0 \leq t \leq t_1, \\
\beta_0 - \alpha_2(t - t_1) - \alpha_3(t - t_1)^p & \text{for } t > t_1, 
\end{cases}$$

where $\alpha_1, \alpha_2, \alpha_3$ are reliability index deterioration rates, $t_1$ is the deterioration initiation time, and $p$ is a parameter related to the nonlinearity effect in terms of a power law in time. An increase in $p$ results in an increase in the rate of reliability index deterioration. Note that the reliability profiles in Eqs (3) and (4) are not computed from state functions as explained in Section 2.2, but simulated by the Monte Carlo method with parameters directly assessed by experts or obtained from statistical data.

**Deterioration model with maintenance.** There are two main types of maintenance actions that could affect the reliability index profile: preventive maintenance and essential maintenance. As indicated in Das[41,42] and Frangopol & Das[43], preventive maintenance actions such as major repairs and replacement of damaged members are normally undertaken when the reliability index is above its target value (minimum acceptable reliability level). In general, preventive maintenance delays the deterioration process and, therefore, reduces the rate of reliability index deterioration over a period of time. Essential maintenance actions such as major repairs and replacement of damaged members are normally undertaken when the reliability of the structure has fallen in the vicinity of the target value. The purpose of essential maintenance is to substantially improve the reliability and condition. In general, deteriorating structural systems experience multiple maintenance interventions during their lifetime.

For some preventive maintenance actions, the time of application is prescribed. For example, painting of a steel bridge every ten years or lubricating its bearings every five years belong to this group. These actions are classified as time-based. Other maintenance actions are applied when a specific performance requirement is, or is close to being violated. For instance, an element should be repaired if it reaches a target reliability level or a bridge component should be replaced if the corrosion penetration and/or spread is too high. These actions are called performance-based (e.g. condition-based or reliability-based) interventions. Time- and performance-based maintenance could be applied once or cyclically during the service life of a deteriorating structure. Time-based maintenance can be combined with performance-based maintenance during a specified time horizon. This time horizon could be limited to a part of the remaining service life of an existing structure or extended to the entire service life of a new structure. The effect of any maintenance action $i$ on the initial reliability index profile $j$ associated with no maintenance can be expressed as a reliability index profile. For example, the reliability index profile associated with maintenance including the effects of all maintenance actions over a given time horizon is obtained by superposition as follows[44]:

$$\beta_i(t) = \beta_{i,0}(t) + \sum_{i=1}^{n_i} \Delta \beta_{i,i}(t),$$

where $n_i$ is the number of maintenance actions associated with reliability index profile $j$, $\beta_{i,0}(t)$ is the reliability index profile without maintenance, and $\Delta \beta_{i,i}(t)$ is the additional reliability index profile associated with the $i$th maintenance action. Finally, the reliability index profile of the system is obtained by statistically combining the reliability index profiles of all individual elements and limit states considered.
Frangopol\cite{45} defined an eight random variable model (see also\cite{46}) associated with every aspect of the preventive and essential maintenance processes. The eight random variables are shown in Fig. 2.

The initial model proposed by Frangopol based on the reliability index was extended to incorporate a condition index (possibly resulting from visual inspections), and their time-dependent interaction. Consequently, the eight random variable time-dependent reliability index model under maintenance previously mentioned was adapted by Frangopol and his co-workers\cite{40,47,48} to consider performance as defined by both condition and reliability. The condition and reliability indices are considered constant for a period equal to the time of damage initiation, \( t_{IC} \) and \( t_1 \), respectively. After that period, a linear deterioration rate is considered for both condition and reliability, defined by the deterioration rates, \( \alpha_c \) and \( \alpha_t \), respectively. If no maintenance is considered, the time-dependent reliability index is defined as shown in eq. (3) and the time-dependent condition index, \( C(t) \) at time \( t \geq 0 \), is defined as follows:

\[
C(t) = \begin{cases} 
C_0 & \text{for } 0 \leq t \leq t_{IC} \\
C_0 - \alpha_c(t - t_{IC}) & \text{for } t > t_{IC}
\end{cases}
\]  

(6)

where \( C_0 \) is the initial condition and \( C(t) \) is the condition at time \( t \) which is assumed to decrease with time. The effects of maintenance actions on the condition are defined by using the superposition method as described in eq. (5). The interaction between condition and reliability is modelled through correlations among the parameters defining both profiles, as well as by deterministic relations between profiles at each time step\cite{48}.

**Decision model.** Maintenance cost is often considered as fixed and independent of both the state of the structure and the effect of a maintenance action on reliability and/or condition of the structure. However, the cost of maintenance depends not only on the type of maintenance action, but also on the reliability and/or condition state of the structure before and after its application. As an example, the cost of repairing a corroded steel girder depends on the degree of corrosion and the extension of the repair. A model describing this interaction has been presented\cite{49,50} where \( c \) is the total cost associated with a maintenance action, \( \Delta \beta \) and \( \Delta C \) are the improvement in reliability and condition indices, respectively, \( t_{ID} \) and \( t_{DC} \) are the delay in deterioration of reliability and condition indices, respectively, \( c_1 \) is the fixed cost, \( c_2 \) and \( c_4 \) are the costs associated with reliability and condition improvement, respectively, \( c_3 \) and \( c_5 \) are costs associated with delay in reliability and condition deterioration, respectively, \( q_1 \) and \( q_2 \) are parameters associated with the relation between maintenance cost and reliability and condition improvement, respectively, and \( r_1 \) and \( r_2 \) are parameters associated with the relation between maintenance cost and delay in reliability and condition deterioration, respectively.

In addition to the effect of the increase in the reliability index and/or condition indices, \( \Delta \beta \) and \( \Delta C \), and the delay in reliability and condition deterioration, \( t_{ID} \) and \( t_{DC} \), the effect of time of application of each maintenance intervention can also be considered, especially, when investment decisions are made for optimal design\cite{51}. In fact, the same amount of money spent in two different instants has different present values. For investment decisions, costs can be compared if converted to the present value at time zero as follows:

\[
c_t = \frac{c_t}{(1 + r/100)^t}.
\]

(8)

where \( c_t \) is the cost at time \( t \) and \( r \) is the annual discount percentage. Fig. 3 qualitatively shows the

---

\( \beta_0 \) = Initial reliability index

\( t_I \) = Time of damage initiation

\( \alpha_c \) = Reliability index deterioration rate without maintenance

\( \gamma \) = Immediate improvement in reliability index after the application of preventive maintenance

\( t_{IP} \) = Time of first application of preventive maintenance

\( t_0 \) = Time of reaplication of preventive maintenance

\( t_{RP} \) = Duration of preventive maintenance effect on bridge reliability

\( \theta \) = Reliability index deterioration rate during preventive maintenance effect

---

*Fig. 2* Reliability profiles and associated random variables for the options with or without preventive maintenance (after\cite{49})

---

Copyright © 2004 John Wiley & Sons, Ltd.
mean and standard deviation of the condition and reliability profiles, along with the mean cumulative maintenance cost of a deteriorating structure under maintenance or no maintenance. The effect of the discount rate of money is also indicated.

Applications. A number of applications and topics relating to the time-dependent reliability profile and/or condition indices are available: updating with visual inspection results\(^{[52,53]}\), bridge deck replacement\(^{[54,55]}\), bridges in general (steel and concrete)\(^{[33,34,38,39,46,56–59]}\), and bridge system reliability\(^{[58,60]}\).

![Fig. 3 Condition, reliability and cost interaction under maintenance and no maintenance](image)

Also, a case study on a bridge stock composed of steel/concrete, reinforced concrete, prestressed concrete and post-tensioned concrete is considered. The data associated with this bridge stock is provided in Maunsell Ltd and the Transport Research Laboratory\(^{[61]}\). The composition and age of the bridge stock is shown in Fig. 4.

In order to obtain the optimal maintenance strategy for each bridge type, the present value of the expected cost of maintenance without and with preventive maintenance was computed. Fig. 5 shows the present value of expected annual maintenance cost in pounds sterling per square metre of steel–concrete (denoted as steel/concrete) composite bridges assuming no preventive maintenance has been done (i.e. essential maintenance only), preventive maintenance has been done (i.e. preventive and essential maintenance), and only preventive maintenance has been done.

As can be seen in Fig. 5, the cost of maintenance based on essential maintenance only, is much more expensive than that based on preventive maintenance, especially considering the effect of the road user delay cost, also called user cost\(^{[62]}\). Considering the age distribution of the entire bridge stock in Fig. 4, computations of expected bridge stock maintenance cost were performed. Fig. 6 shows the expected future maintenance costs for the different types of bridges while preventative maintenance is performed.

### 3 Stochastic-process models

Owing to the usual lack of failure data, a reliability approach solely based on lifetime distributions and their unobservable failure rates is unsatisfactory. It is recommended to model deterioration in terms of a time-dependent stochastic process \(\{X(t), \ t \geq 0\}\) where \(X(t)\) is a random quantity for all \(t \geq 0\).
Deterioration is usually assumed to be a Markov process [1, Chapter 5]. Roughly speaking, a Markov process is a stochastic process with the property that, given the value of $X(t)$, the values of $X(\tau)$, where $\tau > t$, are independent of the values of $X(u)$, $u < t$. That is, the conditional distribution of the future $X(\tau)$, given the present $X(t)$ and the past $X(u)$, $u < t$, is independent of the past. Classes of Markov processes which are useful for modelling stochastic deterioration are discrete-time Markov processes having a finite or countable state space called Markov chains (see Section 3.1) and continuous-time Markov processes with independent increments (see Section 3.2) such as Brownian motion with drift (also called the Gaussian or Wiener process) and the gamma process.

In the following sections, two decision models based on stochastic deterioration processes are presented. First, the Markov decision process model with a cost function which allows for both partial and perfect repair. Next, renewal models which are based on renewals bringing a component or structure back to its original condition (i.e. perfect repair). Also these renewal models can be extended with the possibility of imperfect repairs.

### 3.1 Markov decision processes

**Deterioration model.** Assume that there is a finite or countable state space. This means that the condition of a structure or component can be in any one of $N \geq 0$ discrete states. A Markov chain is a discrete-time stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$ for which the Markovian property holds. As previously mentioned, this property states that the future condition only depends on the current condition. The conditional probability of moving into state $j$ at time $n + 1$ given that at the current time $n$ the object is in state $i$ is given by:

$$P_{ij} = P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1) = P(X_{n+1} = j | X_n = i).$$

If this transition probability does not depend on $n$ (i.e. does not depend on how long the process has been running), then the process is called stationary in time.

To define the Markov chain $X_n$, it is necessary to assess the transition probabilities between all possible condition state pairs. If there are $N$ states, then this results in a $N \times N$ matrix

$$P = \begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1N} \\
P_{21} & P_{22} & \cdots & P_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
P_{N1} & P_{N2} & \cdots & P_{NN}
\end{bmatrix} \tag{9}$$

This is called a one-step transition matrix, because it gives the transition probabilities for one time unit. Obviously, the probability of moving from one state to any other state (including itself) should be 1, or $\sum_{j=1}^{N} P_{ij} = 1$ for $i = 1, \ldots, N$. The probability $P_{ij}^m$ of moving from the current state $i$ to state $j$ in $m$ time steps is:

$$P_{ij}^m = Pr(X_{n+m} = j | X_n = i).$$

It can be shown [63] that the $m$-step transition matrix can be calculated by multiplying the matrix $m$ times with itself: $P^m$.

**Decision model.** In order to be able to make decisions about an optimal policy for maintenance actions, a finite set of actions $A$ and costs $C(i, a)$ have to be introduced, which are incurred when the process is in state $i$ and action $a \in A$ is taken. The costs are assumed to be bounded and a policy is defined to be any rule for choosing actions. When the process is currently in state $i$ and an action $a$ is taken, the process moves into state $j$ with probability

$$P_{ij}(a) = Pr(X_{n+1} = j | X_n = i, a_n = a). \tag{10}$$
This transition probability again does not depend on the state history. If a stationary policy is selected, then this process is called a Markov decision process (MDP). A stationary policy arises when the decision for an action only depends on the current state of the process and not on the time at which the action is performed.

Now that the state of the structure over time with or without performing maintenance actions can be modelled, the optimization of inspection and/or maintenance policies using this process can be performed. For example, when the system is in state \( i \) the expected discounted costs over an unbounded horizon are given by the recurrent relation

\[
V_s(i) = C(i,a) + \alpha \sum_{j=1}^{N} P_{ij}(a) V_s(j),
\]

where \( \alpha \) is the discount factor for one year and \( V_s \) is the value function using \( \alpha \). This discount factor is defined as \( \alpha = (1 + r/100)^{-1} \) with \( r \) the yearly discount percentage as in eq. (8). Starting from state \( i \), \( V_s(i) \) gives us the cost of performing an action \( a \) given by \( C(i,a) \) and adds the expected discounted costs of moving into another state one year later with probability \( P_{ij}(a) \). The discounted costs over an unbounded horizon associated with a start in state \( j \) are given by \( V_s(j) \), therefore eq. (11) is a recursive equation. The choice for the action \( a \) is determined by the maintenance policy and also includes no repair.

A cost-optimal decision can now be found by minimizing eq. (11) with respect to the action \( a \). There are a number of ways to find this optimal solution. One of these is the so-called policy improvement algorithm, where eq. (11) is calculated for increasingly better policies until no more improvement can be made. A policy iteration algorithm operating on simple rules has been described[64], and later corrected[65]. Also, it is possible to formulate the minimization problem as a linear programming problem. This is used in the Arizona pavement management system (see[66] for a general introduction,[67] for more mathematical implementation details and[68] for an update), PONTIS[69–71], and BRIDGIT[72,73]. As Golabi[67] illustrates, we can choose to maximize the condition of the road system under a budget constraint or we can minimize the maintenance cost under a minimum safety constraint. This can be achieved by using the original linear programming formulation or with its dual formulation. The interested reader is referred to textbooks by Ross[63] and Derman[74] for more details on solution algorithms for Markov decision processes. For easy step-by-step algorithms, see Denardo[75]. An application to the deterioration of coating on steel bridges is given in[76], where a policy improvement algorithm is used. Rens et al.[77] discuss the application of the Markov model to bridge maintenance for the city and county of Denver. In this review, we have restricted ourselves to the discrete-time model with fixed time units. However, it is also possible to apply a so-called semi-Markov decision process, which assumes the time steps between transitions to be random. For an application to infrastructure inspection and renewal planning, see e.g.[78].

There are three important issues concerning the use of Markov decision processes for maintenance optimization:

- the condition state is not continuous, but discrete and finite. This works well for visual inspections, but it is not suitable for nondestructive evaluation (NDE) inspection results. The richer continuous measurement data would have to be categorized into discrete states before it can be incorporated into the model. A possible approach to integrating NDE into bridge management systems is available [79];
- the Markovian assumption of no memory has often been criticized. Some attempts at testing time (in)dependence have been made by performing a hypothesis test using available data, see e.g.[80] and[81].
- transition probabilities in the transition matrix (eq. 9) are difficult to assess and quite subjective. Besides the use of expert judgment, several more quantitative approaches have been suggested in the literature.

Concerning the quantitative assessment of transition probabilities, see[82] where New York bridge data are used,[83] using data from Virginia and where the Indiana Bridge Inventory data are used. The latter advocates the relationship of the transition probabilities with the hazard rate function of a bridge. Also, it discusses previous approaches and supplies a good list of references to older publications. Amongst others, it discusses the ordered probit model[85] and the Poisson regression model[84]. Other data-fitting approaches are discussed in[85] and[86]. Generally, all these approaches require adequate amounts of data, which in most cases are not available. This lack of data is also discussed in[87] for the road network in Portugal. A very different approach from all of the previously mentioned, has been discussed[88], where a continuous-time Brownian motion with drift is discretized into a finite number of states. In[89] and[90], the reliability profile is discretized into a finite number of states and the transition matrix is calculated by the same reliability approach. Also, attempts to analyze a bridge at the system level by expanding the component state space to a system state space were made. Approximate solutions for systems of finite state components are available[91].

Finally, it is noted that it is possible to include measurements from imperfect inspections into the
Markov model. An imperfect inspection has a measurement error associated with it. In [92], it is referred to as a latent Markov decision process, but it is more commonly known as a partially observable Markov decision process (POMDP). In this approach, it is assumed that the observed state \( Y_n \) is probabilistically related to the true state \( X_n \) by

\[
q_{ij} = \Pr(Y_n = j | X_n = i).
\]

A suitable probability distribution function can be used to represent the likelihood \( q_{ij} \) of measuring a state \( i \), when the component or structure is actually in state \( j \). More information can be found on POMDPs in the field of operations research. The added uncertainty of partial observability also adds significant computation time. Quite a few algorithms have been proposed for solving POMDPs, which will not be discussed here. One of the best introductions to using POMDP in operations research. The added uncertainty of partial observability can be found on POMDPs in the field of network optimization of both replacement and repair using a finite state Markov model is presented in [98].

### 3.2 Renewal Models

**Deterioration model.** Brownian motion with drift is a stochastic process \( \{X(t), t \geq 0\} \) with independent, real-valued increments and decrements having a normal distribution with mean \( \mu t \) and variance \( \sigma^2 t \) for all \( t \geq 0 \) [98, Chapter 7]. A characteristic feature of Brownian motion with drift—in the context of structural reliability—is that a structure’s resistance would alternately increase and decrease, similar to the exchange value of a share. For this reason, the Brownian motion is inadequate in modelling deterioration which is monotone. In order for the stochastic deterioration process to be monotonic, we can best consider it as a gamma process [99]. In the case of a gamma deterioration, structures can only decrease in strength. To the best of the authors’ knowledge, Abdel-Hameed [100] was the first to propose the gamma process as a proper model for deterioration occurring random in time. The gamma process is suitable to model gradual damage monotonically accumulating over time, such as wear, fatigue, corrosion, crack growth, erosion, consumption, creep, swell, a degrading health index, etc. For the mathematical aspects of gamma processes, see [101–104].

In mathematical terms, the gamma process \( \{X(t), t \geq 0\} \), with shape function \( v(t) > 0 \) and scale parameter \( u > 0 \), is a continuous-time stochastic process with independent gamma-distributed increments. A random quantity \( X \) has a gamma distribution with shape parameter \( v > 0 \) and scale parameter \( u > 0 \) if its probability density function is given by:

\[
Ga(x|v, u) = \frac{u^v}{\Gamma(v)} x^{v-1} \exp\{-ux\}, \quad \text{for } x \geq 0
\]

where \( \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} \, dt \) is the gamma function for \( a > 0 \). Let \( X(t) \) denote the deterioration at time \( t \). If \( X(t) \) is a gamma process with shape function \( v(t) \) and scale parameter \( u(t) \), then its expectation and variance are given by

\[
E[X(t)] = \frac{v(t)}{u(t)}, \quad \text{Var}(X(t)) = \frac{v(t)}{u(t)^2}
\]

A component is said to fail when its deteriorating resistance, denoted by \( R(t) = r_0 - X(t) \), where \( r_0 \) is the initial resistance, drops below the stress or load \( s \). The initial resistance and the applied stress can be assumed fixed (i.e. known) or random. Let the time at which failure occurs be denoted by the lifetime \( T \). Due to the gamma distributed deterioration, the lifetime distribution can then be written as:

\[
F(t) = \Pr[T < t] = \Pr\{X(t) > r_0 - s\}.
\]

Under the assumption of modelling the temporal variability in the deterioration with a gamma process, the question which remains to be answered is how its expected deterioration increases over time. Empirical studies show that the expected deterioration at time \( t \) is often proportional to a power law:

\[
E[X(t)] = \frac{v(t)}{u(t)} = at^b
\]

for some physical constants \( a > 0 \) and \( b > 0 \). There is often engineering knowledge available about the shape of the expected deterioration in terms of the parameter \( b \) in eq. (15). Some examples of expected deterioration according to a power law are the expected degradation of concrete due to corrosion of reinforcement (linear: \( b = 1 \)), sulphate attack (parabolic: \( b = 2 \)), diffusion-controlled ageing (square root: \( b = 0.5 \)), and creep (\( b = 1/2 \)), and the expected scour-hole depth (\( b = 0.4 \)).

As an illustration, corrosion data of five steel gates of the Dutch Haringvliet storm-surge barrier have been studied. The data represent the percentage of the surface of a gate that has been corroded due to ageing of the coating. Every gate was inspected once at a different inspection interval. When we assume the five inspection results to be independent, we can determine the maximum-likelihood estimators of the parameters of the gamma process with expected deterioration being a power law in time. Fig. 7 shows the expected condition with its 5th and 95th percentile including the inspection data, the probability density function of the lifetime, and the cumulative distribution function of the lifetime (the deterioration failure level is a corroded surface of 3%).

The gamma process is called stationary if the deterioration is linear in time, i.e. when \( b = 1 \) in eq. (15), and nonstationary when \( b \neq 1 \).
mathematical terms, a stochastic process has stationary increments if the probability distribution of the increments $X(t + h) - X(t)$ depends only on $h$ for all $t, h \geq 0$. Çinlar et al.\cite{106} use a nonstationary gamma process as a model for the deterioration of concrete due to creep. They give a comprehensive justification of the gamma process from a physical and practical point of view. In doing so, the expected deterioration agrees with a deterministic creep law in the form of a power law in time. Çinlar et al.\cite{106} show how a nonstationary gamma process can be transformed into a stationary gamma process and how the parameters of a stationary gamma process can be estimated using the method of moments and the method of maximum likelihood. They perform a statistical analysis of data on concrete creep.

**Decision model.** The gamma deterioration process was successfully applied to model time-based preventive maintenance as well as condition-based preventive maintenance. Time-based preventive maintenance is carried out at regular intervals of time, whereas condition-based maintenance is carried out at times determined by inspecting or monitoring a structure’s condition.

All maintenance models surveyed in this section use cost-based criteria which are defined over an unbounded time horizon, such as the expected average cost per unit time and the expected discounted cost over an unbounded horizon. Because these cost-based criteria are all based on renewals bringing a component or structure back to its original condition, renewal theory was used to compute them. Let $F(t)$ be the cumulative probability distribution of the time of renewal $T \geq 0$ and let $c(t)$ be the cost associated with a renewal at time $t$. From renewal theory [63, Chapter 3], the expected average cost per unit time is

$$\lim_{t \to \infty} \frac{E(K(t))}{t} = \int_0^\infty c(t) dF(t) \int_0^\infty t dF(t)$$

where $E(K(t))$ represents the expected nondiscounted cost in the bounded time interval $(0, t]$, $t > 0$. Let a renewal cycle be the time period between two renewals, and recognize the numerator of eq. (16) as the expected cycle cost and the denominator as the expected cycle length (mean lifetime). By renewal theory with discounting\cite{108}, the expected discounted cost over an unbounded horizon can be written as

$$c_0 + \lim_{t \to \infty} E(K(t, x)) = c_0 + \int_0^\infty \frac{x' c(t) dF(t)}{1 - \int_0^\infty x' dF(t)}$$

where $c_0$ is the investment cost, $0 < x < 1$ is the discount factor, and $E(K(t, x))$ is the expected discounted cost in the bounded time interval $(0, t]$.
Similar results can be obtained for discrete-time renewal processes \[ t > 0 \].

Models for time- and condition-based maintenance under gamma-process deterioration are summarized next.

**Time-based maintenance.** Abdel-Hameed \[110\] studied an age replacement policy for which a renewal is defined as either a corrective replacement upon failure or upon a preventive replacement reaching a predetermined age \( k \), whichever occurs first. The cost of a corrective replacement is \( c_F \), whereas the cost of a preventive replacement is a function of the amount of deterioration \( X(k) \) being of the form \( c_P(X(k)) \). The long-term expected average cost of replacement per unit time can be calculated by renewal theory \[63, Chapter 3\]. The failure level is assumed to be random and failure is detected immediately. Abdel-Hameed \[110\] studied the stationary gamma process. He also presented the results for the discrete-time case. As a special case he studied the discrete-time version of the ordinary age replacement model (with fixed preventive and corrective replacement cost \( c_P \) and \( c_F \) for which the increments of deterioration are exponentially distributed and the failure level is fixed. Van Noortwijk \[111\] applied this special case to age replacement of a cylinder on a swing bridge. As a cost criterion, he considers the expected discounted cost over an unbounded time horizon.

The age replacement model with fixed failure level has been extended for the possibility of nonstationary gamma process deterioration and lifetime-extending maintenance by van Noortwijk \[111\] and Bakker et al. \[112\]. With the extended model both the interval of lifetime extension and the interval of preventive replacement can be optimized. Lifetime-extending maintenance is also referred to as imperfect or partial repair: the condition of the structure is improved, but not to its initial condition. Through lifetime extension, the deterioration can be delayed so that failure is postponed and the lifetime of a component is extended (e.g., a coating protecting steel). Possible effects of lifetime-extending measures are the increase in the damage-initiation period (time interval in which no deterioration occurs) and the condition improvement. Through replacement, the condition of a component is restored to its original condition. The model has been successfully applied \[112\] to optimize the application of a protective coating on steel (e.g., a bridge deck). In Fig. 8, the lifetime of the steel can be extended by grit blasting (with 0.1 mm loss of steel thickness), as well as placing the new coating. The expected condition (in terms of the surplus in steel thickness) without lifetime extension represents the corrosion process, given initiation at time zero.

**Condition-based maintenance.** Modelling the deterioration as a gamma process is especially suitable when inspections are involved. Over the last decade several inspection models have been developed for the purpose of optimizing condition-based maintenance. In this section, we give an overview of these condition-based maintenance models and their differences and similarities.

Abdel-Hameed \[113,114\] studied an optimal periodic inspection policy model based on the class of increasing pure jump Markov processes. He provides a hypothetical example using the gamma process, which is a member of this class. The two decision variables are the inspection interval and the preventive maintenance level. In the operations research literature, such a policy is called a ‘control-limit policy’ with the preventive maintenance level called the ‘control limit’ \[11\]. The failure level is assumed to be uncertain, having a probability distribution, and a failure is detected only by inspection. As is shown in Fig. 9, the system is

![Fig. 8](https://example.com/fig8.png)

*Fig. 8 Expected condition of a steel bridge deck with lifetime extension in terms of grit blasting and reapplication of coating*
studied by Kong & Park[115]. However, they assumed corrective replacement are assumed to be fixed. The cost of preventive replacement, and the cost of average cost per unit time. The cost of inspection, the determined by minimizing the long-term expected new’ condition. The optimal maintenance decision is A renewal brings the system back to its ‘as good as

failure level

failure has occurred (preventive replacement) or that preventive maintenance level

renewed when an inspection reveals either that the preventive maintenance level \( \rho \) is crossed while no failure has occurred (preventive replacement) or that the failure level \( s \) is crossed (corrective replacement). A renewal brings the system back to its ‘as good as new’ condition. The optimal maintenance decision is determined by minimizing the long-term expected average cost per unit time. The cost of inspection, the cost of preventive replacement, and the cost of corrective replacement are assumed to be fixed.

Abdel-Hameed’s gamma process example was also studied by Kong & Park[115]. However, they assumed failure to be immediately detected without inspection. A special case of[115] is the model of Park[116], who considered the failure level to be fixed rather than random. Jia & Christer[117] proposed using[116] for modelling functional checking procedures in reliability-centred maintenance (RCM). Van Noortwijk et al.[118] approximated the stationary gamma process with a discrete-time stochastic process having exponentially distributed increments. They present a case study on condition-based maintenance of the rock dumping of the Eastern Scheldt barrier for which the expected average cost per unit time or the expected discounted cost over an unbounded time horizon are minimal. The deterioration process considered is current-induced rock displacement near the rock dumping.

Optimal inspection intervals for a steel pressure vessel subject to corrosion for which the expected average costs of inspection and maintenance per unit time are minimal have been determined by Kallen & van Noortwijk[119]. Imperfect inspections have been dealt with by the same authors as well as by Newby & Dagg[120].

Dieulle et al.[121] and Grall et al.[122] studied the following variation of the inspection model of Park[116]. Inspection is aperiodic and it is scheduled by means of a function of the deterioration state. Failure is detected only by inspection and a cost of ‘inactivity of the system’ per unit time is incurred as soon as failure occurs. Grall et al.[123] approximated the stationary gamma process with a discrete-time stochastic process having independent, identically and exponentially distributed increments. They consider a maintenance policy using a multi-level control-limit rule, where failures are detected immediately. Recently, Castanier et al.[124] extended this discrete-time model with the possibility of partial repair.

Van Noortwijk & Klatter[107] developed a mathematical model, which includes the use of a nonstationary gamma process, to optimize maintenance of a part of the seabed protection of the Eastern Scheldt barrier, namely the block mats. This model enables optimal inspection decisions to be determined on the basis of the uncertainties in the process of occurrence of scour holes and, given that a scour hole has occurred, of the process of current-induced scour erosion. The model of van Noortwijk & van Gelder[125] studies berm breakwaters under uncertain rock transport caused by extreme waves. It computes inspection intervals having either minimal expected cost per unit time or minimal expected discounted cost over an unbounded horizon. Both models include the cost of inspection, repair and failure; the cost of repair depends on the amount of deterioration. In the former model failure is detected only by inspection, whereas in the latter failure is detected immediately.

Apart from using inspection models under gamma process deterioration in the application phase of the life-cycle of a structure, they can also be used in the design phase. Only two inspection models have been devoted to optimally balancing the initial investment cost against the future maintenance cost in the design phase. The first model deals with determining optimal sand nourishment sizes in which the stochastic process of permanent coastal erosion of dunes is regarded as a stationary gamma process[51]. The second model deals with determining optimal dike heightenings in which the stochastic process of crest-level decline is regarded as a stationary gamma process[126]. Both models use a discrete timescale and assume that maintenance costs depend on the amount of deterioration and are discounted. In the former model failure is detected immediately (at discrete units of time), whereas in the latter model failure is detected only by inspection.

4 Future directions

An attempt has been made to provide the reader with a review of probabilistic models for maintaining and optimizing the life-cycle performance of deteriorating structures, with the emphasis on bridges.

A variety of different modelling approaches have been discussed. Some primarily deal with the
reliability index, whereas others are concerned with the physical condition of a structure. No single approach has yet proven to be generally applicable and each model has its advantages and disadvantages. For example, the Markov decision model is purely a condition model and is very well suited to incorporate information from visual inspections, but it cannot be used to assess the reliability of a structure in terms of strengths and stresses.

Currently, the Markov model is the most commonly used approach in bridge maintenance models. The use of the reliability index to model the performance of a structure is a classic approach in civil engineering and has resulted in many design codes based on this index. The gamma process model has been the subject of many scientific publications with a few applications to real maintenance problems in civil engineering. In the Netherlands, an effort has been made to apply these models to the maintenance of bridges on the national road network.

Further work is necessary to collect relevant data, improve the modelling capability and formulate probabilistic decision problems as follows:

- establish a general acceptable and consistent methodology for probabilistic modelling of deterioration processes of structural performance in terms of both condition and reliability;
- improve the understanding of the effects of maintenance actions on structural performance and their probabilistic modelling, improve the incorporation of measurement data from imperfect inspections into the deterioration models;
- develop consistent probabilistic methodologies for evaluating maintenance and management strategies;
- use optimization for finding the best strategy through balancing of competing objectives such as reliability, condition, and cost.

Acknowledgements

The work reported in this paper was partly supported by the Civil Engineering Division of the Netherlands Ministry of Transport, Public Works, and Water Management, by the US National Science Foundation through grants CMS-9506435, CMS-9912525, CMS-0217290, and by the UK Highways Agency. This support is gratefully acknowledged.

The opinions and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the sponsoring organizations.

References


