A model for determining strategies to maintain single components of a system at minimal costs

D.J.D. WIJNMALEN and J.A.M. HONTELEZ
TNO Physics and Electronics Laboratory
Division of Operational Research

Abstract

In this paper we present a model which determines strategies to maintain single components of a system at minimal (discounted) costs. We focus on civil-technical components which deteriorate gradually. Information about a component's condition is obtained by inspection only, except for an obvious breakdown. Repair returns the component to the as-good-as-new condition. Decisions have to be made on when to inspect (inspection intervals may depend on conditions found) and at which condition level to repair. The problem is modelled as a semi-Markov decision process. A special-purpose iteration procedure leads to an optimal strategy belonging to the class of control-limit rules. This type of maintenance strategy is useful for practical purposes.
We discuss some results obtained with the model and we indicate some extensions.

Keywords: Maintenance strategy, inspection, repair (replacement), optimisation, semi-Markov decision process, deterioration process.

1 Introduction

In another paper of this issue, Van der Toorn points out that several compelling reasons exist for developing a sound theoretical basis for maintenance planning. Especially, but by no means exclusively, where government expenditure is involved and very large amounts of money go into maintaining public works and where at the same time budgets are getting tighter, careful (re)consideration of maintenance strategies are called for. Effective and efficient maintenance management and control involve analysing and balancing maintenance costs and risks of failure related to preventive and corrective actions. This paper deals with the development of maintenance strategies for an equipment or a structure consisting of components which deteriorate gradually. If costs of failure (or malfunctioning) are very high, preventive maintenance may be called for. Preventive maintenance actions consist of inspection in order to reveal the exact working condition of the equipment or the actual degree of deterioration if there is no continuous monitoring of the system, and of repair or replacement if the actual condition is equal to or exceeds a pre-defined condition level. This level is called "repair limit". Decisions have to be made as to the length of inspection intervals and the repair limit. The result is a maintenance strategy which minimises total maintenance costs consisting of cost of failure, operation, inspection and repair.

The model we are going to present is based on a model by Tijms & Van der Duyn Schouten (1984) and improved and extended further to meet specific requirements from practice by Hontelez & Wijnmalen (1991), Wijnmalen (1992), and Wijnmalen & Hontelez (1992). The model was implemented in a PC-software package called OPTIMON and has been in actual use by the Netherlands Department of Public Works for some time now.

2 The model

In this section a description of the model will be given. At first, we focus on the general model formulation for determination of condition-based strategies that minimise discounted costs. In a later subsection we shall indicate how other basic maintenance concepts can be derived from this model.

2.1 *Assumptions*

We focus on structures or equipment which perform specific functions on a permanent basis and which consist of one or more components. One might think of bridges with components made of steel, concrete, etc., and mechanical components, all of which may be protected by a coating (e.g. paint). We investigate each component individually; we do not take any economic or technical dependencies between components into account (see, however, section 4). We do not take two-or-more-layer situations into account either in this paper.

Processes like corrosion, carbonation, wear, and shrinkage cause a gradual deterioration of the civil-technical components. We take these kinds of deterioration processes into account, but not failures of electro-technical components. We assume that a component can be observed in either one of N discrete conditions $i = 1, ..., N$, where $i = 1$ stands for the condition "new" and $i = N$ stands for failure or malfunction. The condition is thus described using one single parameter. The number of conditions can be chosen at will, but should reflect the nature of the actual deterioration process, the accuracy of visual or technical measurement to be achieved, the required level of detail, etc. Failure may be revealed at the next inspection with probability $1\text{-}q_1$ (e.g. the tread of a tyre worn off) or may be detected immediately with probability q_1 (e.g. a burst tyre). Even if the component fails without being noticed before inspection, there may be a possibility of the failure revealing itself after all: we define q_2 as the probability of this event per time unit. In order to make a distinction between the two possibilities of failure detection, we define N as the failure condition detected by inspection, $N + 1$ as the failure condition revealed by itself, and $F = \{N, N + 1\}$ as the state of failure.

Information on the nature and stochasticity of deterioration processes may be given as a mathematical function $G(t)$ which decreases or increases continuously, where $G(t)$ is the condition at time t. $G(t)$ follows a Normal distribution with mean $g(t)$ and variance b^2t:

$$G(t) = g(t) + b*U*\sqrt{t}, \text{ with } U \sim \mathbf{N}(0,1)$$

The mathematical definition of this function depends on the nature of the deterioration process. Making an independence assumption and adopting discrete condition levels, we transform $G(t)$ to a discrete deterioration process, defining $r_{ij}(t)$ as the probability that the component deteriorates during t units of time from a known condition i to condition j $(1 \leq i \leq j \leq F)$. We assume that a component's condition cannot improve on its own. We refer to Burger et al. (1995) or Hontelez & Wijnmalen (1991) for details. In Figure 1 an example is shown where at discrete points of time the condition of a component is assigned to the proper condition interval.

Maintenance actions are inspection, involving costs $CI \geq 0$ per inspection and taking time $TI \geq 0$, and repair (revision, replacement) resulting into the condition $i = 1$ ("new"), involving costs $CR(i) \geq 0$ per repair and taking time $TR(i) \geq 0$. Costs and time of repair can depend on the condition the component is in when repair starts. There may be a delay between the end of inspection and the beginning of a consecutive repair: $TD(i) \geq 0$, where i is the condition detected by inspection. Inspection reveals the exact condition with certainty; uncertainty about the actual condition starts to increase as the deterioration process continues. We divide the (infinite) planning horizon into planning intervals of equal length; one time unit coincides with one planning interval. Opportunities for inspection and repair occur at discrete points of time $t = 0, 1, 2, \ldots$ which are equi-distant. During an inspection or repair, which take an integral number of time units, no other action can be decided on. When no maintenance action is performed, the component works or functions at cost $CO(i) \geq 0$ per time unit, where i stands for the actual (but possibly unknown!) working condition. Upon detection of failure, damage costs $CF \geq 0$ are accounted for (once per failure); per unit of time the component is in a failure state, costs $CO(N) = CO(N + 1) \geq 0$ will occur. All costs can be discounted at rate $\alpha = 1/(1 + r)$, with $0 \leq \alpha < 1$ and r the interest rate.

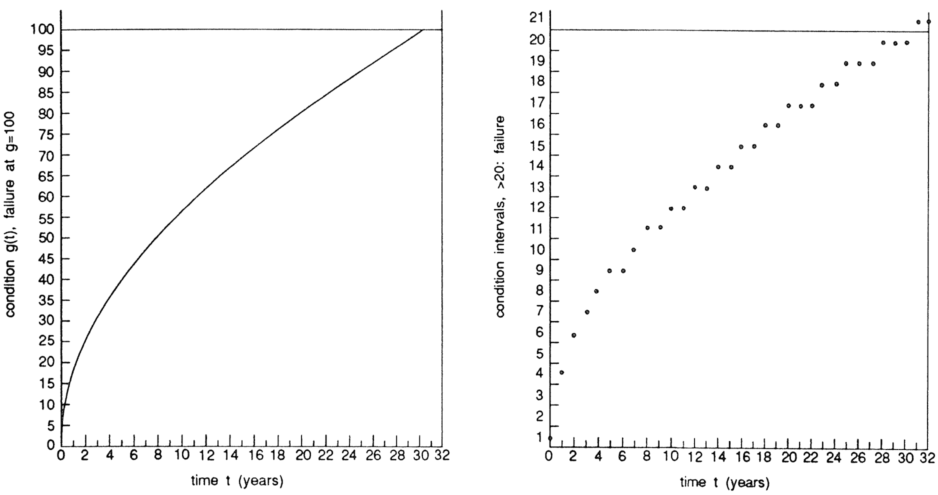

Fig. 1. Discretisation of a continuous deterioration process (Representation of a square-root deterioration process: e.g. carbonation).

The problem characterised above will be modelled as a discrete, semi-Markov decision process, where decisions have to be made as to when to inspect and at which detected condition level to repair. The process is discrete because we shall intervene at specific points of time only. The process is called Markov because the transitions between the states of the decision process depend only on the current state where the decision is taken and on the decision itself. The process is called semi-Markov as the time intervals between decision moments are not of equal length due to varying inspection, delay and repair times.

The maintenance strategies which are considered here, each consisting of a scheme of condition-based inspection intervals and a repair limit, belong to the class of control-limit rules. These rules are simple to implement and thus of practical interest. A control-limit rule R consists of an integer π_0, indicating that repair must be performed if a condition $i \geq \pi_0$ is revealed and not if a condition i with $1 \leq i < \pi_0$ is revealed, and of integers $\pi(1), ..., \pi(\pi_0 - 1)$, indicating that inspection must be performed when $\pi(i)$ time units have passed since the condition was last known to be i. Limits $\pi(j)$ with $j \geq \pi_0$ do not exist as in such conditions j the decision to repair is prescribed and waiting until the next inspection is not allowed. In the failure condition $F = \{N, N+1\}$, repair is always mandatory. Tijms & Van der Duyn Schouten (1984) conjecture that under fairly general conditions this class of control-limit rules will contain the overall cost-optimal strategy. Figure 2 shows an example of a control-limit repair rule, and a rule which is not of the control-limit type; a repair in condition 3 would suggest repairing in condition 4 as well, under a control-limit rule.

condition		action
5 (fail)	⊗	repair
4	⊗	repair
3	⊗	repair
2	O	no repair
1 (new)	O	no repair

control-limit rule (π_0 = 3)

condition		action
5 (fail)	⊗	repair
4	O	no repair
3	⊗	repair
2	O	no repair
1 (new)	O	no repair

not a control-limit rule

Fig. 2. Example of a proper control-limit rule and of a rule which is not of the control-limit type.

We consider stationary strategies only, which means that the limit values do not depend on actual time. No matter at what time a condition i is detected, the decision to repair or to wait π_i time units until the next inspection remains the same, according to the strategy scheme. The consequence is that these strategies can help setting general standards and norms for maintenance, and can thus offer a guideline to actual maintenance. In practice, deviations from such schemes are fairly common and should be taken into account in day-to-day planning procedures. Such planning and control activities are, however, not considered in this paper.

The criterion for optimality is based on a cost function that includes all cost categories mentioned above. We are searching for a strategy (of the control-limit type) which minimises total costs. If costs are discounted, expected total discounted costs are minimised over an infinite horizon; if costs are not discounted ($\alpha = 1$), average costs per unit of time are minimised over an infinite horizon. In this paper we present the discounted costs model; in Burger et al. (1995), Hontelez & Wijnmalen (1991), and Wijnmalen & Hontelez (1992) the undiscounted costs model can be found.
A more formal mathematical description of the model is given in the following subsection.

2.2 *Model formulation*

The model and solution procedure are based on a discrete, semi-Markov decision process formulation (see, for example, Tijms (1986) for a general theoretical introduction).
The principle idea is that we do not follow the whole life-time of a component and investigate all possible sequences and occurrences of events that might take place (leading to unmanageably large event trees), but just list all those possible states of the maintenance process that are relevant to the decision process. In each state we consider those actions that are allowed and which we have to decide upon and investigate the consequences of each action (up to and including the transition to the next state). For each state the optimal decision is determined. As the action chosen results in a transition to one of the states listed (the list should be exhaustive), the new starting point is comparable with the previous starting point.
The state space is defined by:

$$S = \{ (i, m) \mid 1 \le i \le N + 1;\ 0 \le m \le M_i;\ M_N = M_{N+1} = 0 \}$$

where a state (i, m) corresponds to the situation of m units of time having passed (with maximum of M_i) since the last knowledge of the component's condition i. Note that (i, m) is a state of the maintenance decision process and is defined by two parameters: the condition parameter i of the component to be maintained and a time parameter m. This definition allows us to formulate the decision process as a (semi-)Markov process.
Possible actions are denoted by:

$a =$
- 0: leave the component as it is,
  this action is allowed in states of the set
  $S_0 = \{ (i, m) \mid 1 \le i \le N-1;\ 0 \le m \le M_i - 1 \}$
- 1: inspect the component
  this action is allowed in states of the set
  $S_1 = \{ (i, m) \mid 1 \le i \le N-1;\ 1 \le m \le M_i \}$
- 2: repair (revise, replace) the component,
  this action is allowed in states of the set
  $S_2 = \{ (i, 0) \mid 2 \le i \le N + 1 \}$

Note that in each state at most two actions are allowed, and that the decision to repair is allowed only after an immediately preceding inspection (with the one exception of a failure having revealed itself without inspection: in state $(N + 1,0)$). The decision to repair is mandatory if failure is detected and is not allowed in the new-condition. In order to restrict the size of the model, quantities M_i are introduced: they set a limit on the length of the inspection intervals. Should the optimal strategy indicate to wait M_i units of time, then one would be wise to increase the value of this particular M_i in order to investigate if the computed strategy might still be improved. A particular control-limit strategy R prescribes for each state (i, m) the value of the action a, as illustrated by the following example.

Suppose, we have five condition levels $i = 1, \ldots, F = \{5,6\}$, and we impose a maximum inspection interval of $M_i = 4$ for $i = 1, 2, 3, 4$. The list of all relevant states in the model is:

(1,0) (2,0) (3,0) (4,0) (5,0) (6,0)
(1,1) (2,1) (3,1) (4,1)
(1,2) (2,2) (3,2) (4,2)
(1,3) (2,3) (3,3) (4,3)
(1,4) (2,4) (3,4) (4,4)

The strategy $R = (\pi_0 = 3; \pi_1 = 4, \pi_2 = 2)$, which is a control-limit rule, indicates that the next inspection should be carried out after 4 time units if the last known condition is $i = 1$, and after 2 time units if the last known condition is $i = 2$, and that immediate repair (boiling down to replacement in our model) should be carried out if inspection reveals a condition of $i = 3$ or worse. This implies that for each relevant state the value of the action a is known. In the states (1,0) . . . (1,3), (2,0), (2,1) $a = 0$, whereas in the states (1,4) and (2,2) $a = 1$, and in the states (3,0), (4,0), (5,0), (6,0) $a = 2$. Under this particular control-limit rule, all other states will not be attainable, and it does, therefore, not matter what value a has in those states. If we would begin the process in a non-attainable state, then we should take that decision that would bring the process as quickly as possible into the optimal scheme; in state (3,1) we should take $a = 2$ (repair), for example.

An action a according to control-limit rule R in a state s involves the process taking one transition step to a new state s'. There may be several candidate states, each with its own probability value. The decision to inspect in state (1,4), for example, may reveal condition level 1, 2, 3, 4 or 5. Hence in the model, transitions are possible from state (1,4) to states (1,0), (2,0), (3,0), (4,0), and (5,0) if we take action $a = 1$ in state (1,4).

We denote with a_R that action a is taken according to a rule R. Transitions are defined by:

$\overset{(t)}{P}\{s \to s' \mid R\}$:	probability of a transition from state s to state s' in t steps adhering to rule R in all steps, with $s, s' \in S$
$P\{s \to s' \mid a_R\}$:	probability of a one-step transition from state s to state s' taking an action in s according to rule R, with $s, s' \in S$
$C_s(a_R)$:	expected transition costs, made up of costs of inspection, repair, operation and failure; the expectation is taken over all possible transitions out of s under rule R

$T_s(a_R)$: expected transition time if action $a_R = 1$ or $a_R = 2$ is taken; expectation taken over all possible transitions

$T_s(0) = 1$: expected transition time if action $a_R = 0$ is taken; expectation taken over all possible transitions.

Basic formulas can be found in the appendix. The transition probabilities are a function of the deterioration probabilities $r_{ij}(t)$ and the special failure detection probabilities q_1 and q_2.

The problem is now to find optimum values for the parameters of the control-limit rule $R = \{\pi_0; \pi(1), ..., \pi(\pi_0 - 1)\}$ so as to minimise expected total discounted costs in the long run:

$$v_s(R) = \sum_{t=0}^{\infty} \alpha^t \cdot \sum_{s' \in S} P\{s \to s'|R\}^{(t)} \cdot C_{s'}(a_R) \quad (1)$$

for each initial state s. We denote the total expected discounted costs by $v_s(R)$ when starting in a state $s \in S$ and assuming that in this and each subsequent state we take an action according to the rule R. The optimisation is such that, no matter which initial state we are in, adherence to the optimal rule yields minimum costs in the long run.
Note that there is a scheme of costs: to each state pertains a cost value. Normally, one is dealing with state (1,0) (new-condition) when comparing strategies and deciding on maintenance concepts. If one is interfering, however, in an ongoing process, the cost value pertaining to the current state should be taken as reference.

2.3 *Solution procedure*

The solution to the problem described in subsection 2.2 can be obtained using an iteration procedure. This procedure starts with a given strategy R_0, which should be a control-limit rule, and attempts to improve this strategy. Each iteration step consists of two substeps which will be described below, and produces an improved strategy. The procedure ends (after a finite and, usually, limited number of steps) when a strategy cannot be improved; this strategy is the optimum one. For the general Markov-decision theory behind this solution procedure, we refer to Tijms (1986).
The *first substep* of each iteration is concerned with setting up and then solving a set of linear equations (2) in $v_s(R)$ (see subsection 2.2). All $v_s(R)$ are minimised, including $v_{(1,0)}(R)$ which represents the total expected discounted costs starting with the new-condition.

$$v_s(R) = C_s(a_R) + \sum_{s' \in S} \alpha^{T_s(a_R)} \cdot P\{s \to s'|a_R\} \cdot v_{s'}(R), \text{ for all } s \in S \quad (2)$$

Note the conformity between (1) and (2). From (2) we see that expected total discounted costs $v_s(R)$, when starting in a state s and taking action a_R in that particular state, equals the immediate expected costs $C_s(a_R)$ of this action plus the expected total discounted costs from the subsequent state. As there may be several subsequent states which can be reached from s when taking action a_R, we have to take the probability of each possible subsequent state into account. As we are discounting, we have to take the discount rate and the expected transition time into account as well. We have already pointed out in the beginning of subsection 2.2 that we are not considering the whole life-time of a component, but rather one arbitrary transition step from each possible state in the long run.

After straightforward calculations, the above set reduces to the embedded set:

$$v_{(i,0)}(R) = A_i + B_i \cdot v_{(1,0)}(a_R) + \sum_{j=i}^{N+1} C_{ij} \cdot v_{(j,0)}(a_R)\,, \text{ for } i = 1, \ldots, \pi_0 - 1 \quad (3)$$

where A_i, B_i and C_{ij} are constant under rule R. This set (3) is of considerable less size than the set (2) and can be solved using standard methods. The values of the quantities $v_s(R)$ not appearing in the set (3) can then be derived from (2) by single-pass calculations. The equation set (2) and the constants A_i, B_i and C_{ij} of (3) are specified in the appendix.
The *second substep* is concerned with improving the strategy R. In order to do so, a test is performed whether or not increasing or decreasing the limit values of the rule R lead to a test quantity value $T_R(s;a)$ lower than the value of $v_s(R)$, where

$$T_R(s;a) = C_s(a) + \alpha^{T_s(a)} \cdot \sum_{s' \in S} P\{s \to s' | a\} \cdot v_{s'}(R)\,, \quad (4)$$

The definition of this test quantity is very similar to that of $v_s(R)$ in (2). The difference is that this test quantity gives the expected total discounted costs if we would take only once another action a than the prescribed a_R in state s, and from the subsequent state s' on we would again follow the official rule R. If the test quantity value of the alternative action is lower than $v_s(R)$ (note: we are *minimising* costs), the alternative action is taken because it apparently involves less costs. In each state s where the current action is a_R and where an alternative action a is allowed, the result of the test should indicate whether or not this alternative action leads to less costs.
More specifically, the improvement procedure consists of:

a. decreasing (if possible) the repair limit value π_0, and investigating if the test quantity value in states $(i, 0)$ with $2 \le i < \pi_0$ when taking action $a = 2$ is less than the value of taking action $a = 0$ in these states; if so, then go to c, otherwise go to b;
b. increasing (if possible) the repair limit value π_0, and investigating if the test quantity value in states $(i, 0)$ with $\pi_0 \le i < N$ when taking action $a = 0$ is less than the value of taking action $a = 2$ in these states; go to c;
c. decreasing (if possible), successively, each inspection limit value $\pi(i)$ with $1 \le i < \pi_0^{\text{new}}$, and investigating if the test quantity value in states (i, m) with $1 \le m < \pi(i)$ when taking action $a = 1$ is less than the value of taking action $a = 0$ in these states; if not, then go to d;

d. increasing (if possible), successively, each inspection limit value $\pi(i)$ with $1 \leq i < \pi_0^{\text{new}}$, and investigating if the test quantity value in states (i, m) with $\pi(i) \leq m < M_i$ when taking action $a = 0$ is less than the value of taking action $a = 1$ in these states.

We refer to Wijnmalen & Hontelez (1992) for details of the improvement step. If the strategy could not be improved by changing the limit values, we have found the optimum strategy. Otherwise, we compute the new values of $v_s(a_R^{\text{new}})$ by solving (3) in a next iteration step.

2.4 *Performance indicators*

In addition to information on the costs of the strategy computed, other "performance" indicators can be computed, such as:

- expected length of the repair cycle,
- expected availability during the repair cycle,
- expected time of being available during the repair cycle,
- expected life time,
- the expected number of inspections during the repair cycle,
- probability of failure during a repair cycle,

where the repair cycle is defined as the time period between the ends of two successive repairs (replacements).

The computations are fairly simple, but will not be shown here due to space limitations.

2.5 *Modelling different maintenance concepts*

Thus far we have considered the general situation of condition-based inspection intervals and repair limits. Other well-known maintenance concepts are: condition-based maintenance without inspection ("repair model" with perfect information), age- or use-based maintenance and failure-based maintenance. These concepts can be considered as special cases of the general model. We shall indicate how we derive appropriate model formulas.

The condition-based maintenance model without inspection can be obtained by putting *CI* and *TI* equal to zero and M_i equal to 1 for all i with $1 \leq i \leq N - 1$. As a consequence, only states $(i, 0)$ remain and π_0 is the only parameter to be optimised.

The age- or use-based maintenance model can be obtained by putting *CI* and *TI* equal to zero, allowing decision $a = 2$ in state (1,0) and fixing π_0 to 1. This means that inspection has been eliminated and repair actions are taken independently of condition. It can easily be verified that only states $(1, m)$ with $0 \leq m \leq M_1$ and $(i, 0)$ with $2 \leq i \leq N + 1$ remain. The only parameter to be optimised is π_1, which is considered to be the moment of replacement. The set of equations (3) can be reduced to an explicit analytical formula for $v_{(1,0)}(R)$:

$$v_{(1,0)}(R) = \frac{A + B}{1 - C} \tag{5}$$

where A represents discounted operating costs (possibly including expected costs of unnoticed failure), B represents discounted repair costs (possibly including damage costs), and C is a function of α and the expected length of the repair cycle. Basically, formula (5) is the sum of an infinite geometrical progression with ratio C being the discount factor for one full repair cycle and with $A + B$ representing the expected costs of that repair cycle discounted to the starting point of the cycle. We refer to Wijnmalen (1992).

3 Results

In this section we shall present and discuss some model results. We shall focus on a fictitious component of an unknown system.

The input data can be summarised as follows (for notation we refer to sections 2 and 6):

- the value of the condition parameter ranges from 0 (new) to 100 (failure)
- $N = 21$, which implies that there are 20 operational condition levels (intervals of equal length); the failure condition leads to failure states (21,0) (failure detection by inspection) and (22,0) (failure detection without inspection)
- $G(t) = 18\sqrt{t} + 4U\sqrt{t}$ (carbonation process), with t expressed in years and $U \sim N(0,1)$
- $q_1 = 1.0$, $q_2 = 0.0$, which implies that failure reveals itself upon occurrence (i.e. at the next time step of the model horizon)
- $CI = 2000$ (inspection costs)
  $CF = 100000$ (fixed damage costs)
  $CR(i) = 10000$ for all i (repair/replacement costs)
  $CO(i) = 250$ per year (operating costs)
  for $i = 1, \ldots, N-1$
  $CO(N) = CO(N+1) = 5000$ per year (time-dependent damage costs)
- $TI = TR(i) = TD(i) = 0$ for all i (durations)

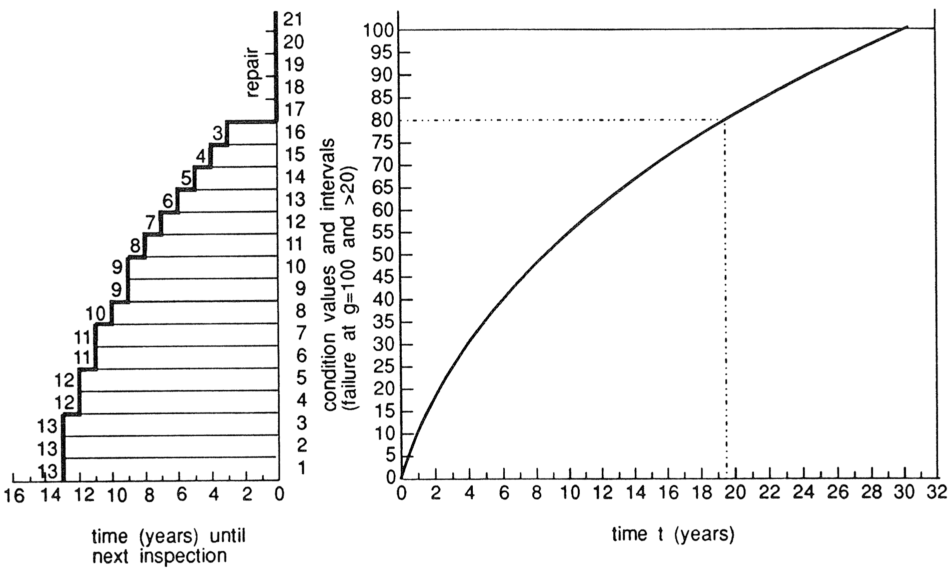

Fig. 3. Optimal condition-based inspection and repair strategy for a carbonation process (see table 1, part 1).

Figure 1 shows the average deterioration process of this example and its discretisation. Table 1, part 1, shows the model results of various maintenance concepts assuming interest rate 5% (discount rate $\alpha = 1/1.05 = 0.952381$). “Total costs” are the total discounted costs $v_{(1,0)}(R)$ over an infinite horizon when starting in the condition “new”. From this table we conclude that in this example the optimal condition-based maintenance strategy is optimal over the other concepts considered. Figure 3 illustrates the relation between this strategy and the (theoretical) deterioration process.

Table 1. Results pertaining to different maintenance concepts (F=failure-based, A=age-/use-based, C=condition-based) and strategies.

concept	total exp. discounted costs (5%)	exp. life time (in years)	time to replace (in years)	repair limit (cond. level)	inspection limits (in years)	comments
F	37119.90	34.3	–	–	–	
A	17034.65	14.0	14	–	–	optimal
C	15738.33	22.0	–	17	13, 13, 13, 12, 12, 11, 11, 10, 9, 9, 8, 7, 6, 5, 4, 3	optimal
C	19271.94	21.7	–	17	5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 3	“common practice”
A	17034.65	14.0	14	–	–	strat. optimal at $r = 0\%$
C	15835.22	20.5	–	16	13, 12, 12, 12, 12, 11, 11, 10, 9, 8, 8, 7, 6, 5, 4	strat. optimal at $r = 0\%$ (N.B. costs are now discounted at 5%)
A	12329.55	21.4	22	–	–	optimal, CF=0 and $q_1 = q_2 = 0$
C	11945.85	28.2	–	19	19, 19, 19, 19, 18, 18, 17, 16, 15, 14, 13, 12, 11, 10, 8, 7, 5, 4	
A	15159.00	15.95	16	–	–	optimal, CF=0 and $q_1 = q_2 = 0$ and CO(N) = CO(N + 1) = 30000
C	13818.84	24.5	–	18	14, 14, 14, 14, 13, 13, 12, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3	

Based on the optimal condition-based strategy, Figure 4 shows how this strategy may be used in practice. The length of the first inspection interval is 13 years. After 13 years, inspection reveals working condition 15, which is worse than the theoretically expected condition 13. From the strategy scheme we conclude that the next inspection should be conducted after 4 years. The actual deterioration will continue. The second inspection then reveals working condition 17, which implies a repair action according to the strategy. Furthermore, the uncertainty as to the actual condition after 13 years, and, again, after 4 years is shown. The shape of the second distribution (at $t = 17$) is due to the fact that the condition observed at $t = 13$ cannot improve of its own during the following 4 years.

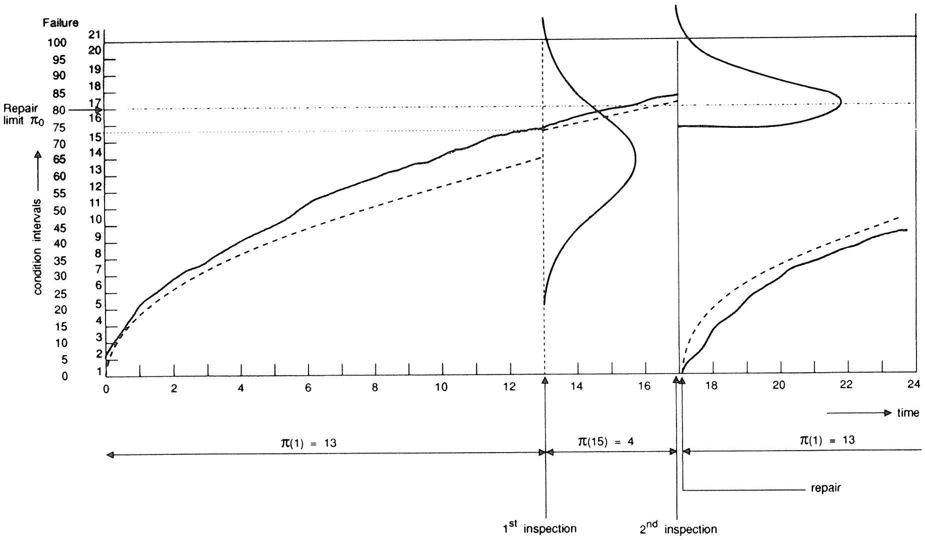

Fig. 4. Application of the optimal condition-based inspection and repair strategy (see table 1, part 1), with difference between assumed and actual deterioration.

As inspection costs are rather high compared to replacement costs, one might expect an age-based maintenance strategy without inspection to be globally optimal. Other effects (such as discounting future costs, reduced uncertainty through inspections about the actual condition and the risk of failure) apparently outweigh this and are the cause that a condition-based strategy with inspection is globally optimal. We also conclude from part 2 of table 1 that a "common practice" condition-based strategy with periodic inspection intervals of equal length (except in bad conditions) is rather expensive.

When we compare the optimal strategies pertaining to interest rate 5% (part 1 of table 1) with those from the undiscounted costs model (part 3 of table 1), we conclude that the optimal strategies differ little or nothing at all. We have calculated the discounted costs (at interest rate 5%) of the optimal condition-based undiscounted costs strategy, as shown in the table; the difference is less than 1%. The higher the interest rate the larger, however, the difference (results not shown).

As for expected life-time, the optimal condition-based strategy combines lower costs with longer expected life than the optimal age-/use-based strategy. Obviously, the failure-based strategy combines the longest life time with the highest cost. The latter is due to the large amount of damage costs.

Part 4 of table 1 shows the results if we take $CF = 0$ (zero fixed damage costs) and $q_1 = q_2 = 0$ (failure is revealed by inspection or at replacement only). In this case of course, the failure-based concept is not feasible. In comparison with the optimal strategies pertaining to the original data, we observe that the optimal age-/use- and condition-based strategies suggest postponing actions: with zero fixed damage costs and relatively low time-dependent failure costs, the occurrence of failure has become less serious. The increase of time-dependent failure costs involves taking action sooner (as shown in part 4 of table 1).

Table 1. Total expected discounted costs (at rate 5%) if the process starts in either one of the states shown.

state $(i,0)$	total expected discounted costs	state $(i,0)$	total expected discounted costs
(1,0) *new*	15738.33	(12,0)	20662.76
(2,0)	15804.33	(13,0)	21691.01
(3,0)	15941.15	(14,0)	22830.30
(4,0)	16133.12	(15,0)	24056.53
(5,0)	16394.16	(16,0)	25306.37
(6,0)	16746.75	(17,0)	25738.33
(7,0)	17154.09	(18,0)	25738.33
(8,0)	17659.77	(19,0)	25738.33
(9,0)	18271.93	(20,0)	25738.33
(10,0)	18959.15	(21,0) *fail*	125738.33
(11,0)	19752.31	(22,0) *fail*	125738.33

Table 2 shows the total expected discounted costs when starting the process in either one of the states $(i, 0)$ and acting according to the optimal condition-based strategy (original data). We observe that the worse the condition is the higher the cost value, as one would expect from the moment of replacement (which involves costs) becoming more imminent as the condition worsens. In states (17,0)...(20,0) the costs are equal: these are the states where the component will be replaced without distinction. In states (21,0) and (22,0) the component will be replaced as well, but an extra charge of 100000 is made due to failure (damage costs).

The slope of the optimum inspection strategy reflects nicely the slope of the deterioration process, as shown in Figure 4. If the component is relatively new, the inspection intervals are relatively long and, consequently, we end up with bad conditions as inspection result. The decreasing worsening rate of condition in the range of relatively bad conditions is thus reflected in the rate of decrease of inspection interval length.

4 Extensions

An obvious extension would be to consider a system of two or more components as a whole, and take into account the effect of savings when maintenance actions on different components are combined. One might think of set-up costs being charged only once. Technical interdependencies are still not considered then.

Theoretically, the general model of section 2 could be extended in a straightforward way by defining a Cartesian product of the state space. The size of the state space, however, would increase exponentially with the number of components, and the set of equations to be solved in the first substep of the iteration procedure would become too large. Therefore, we have developed a heuristic procedure based on aggregation and decomposition. We refer to Wijnmalen & Hontelez (1993) for theoretical details.

Other extensions are:
Making forecasts of maintenance actions and costs over a finite horizon if a fixed strategy (possibly the optimum one) was adopted. This would involve successsive multiplications of the transition matrix. *Optimising* over a finite horizon, however, would require a different model formulation. We are currently developing such a model.
Repair proper besides replacement could be modelled by defining additional decisions in states. Transitions following from such decisions would result in a better condition but not necessarily in the new-condition. By adding an additional state parameter one should be able to control the number of repairs.

5 Conclusions

We have shown how a maintenance process can be modelled as a discrete, semi-Markov decision process in order to minimise total discounted maintenance costs. We have considered single components which deteriorate independently of one another, and have demonstrated that optimum inspection intervals and repair limits can be determined assuming that a component's condition is measurable. We have adopted maintenance strategies from the class of control-limit rules, which are of practical interest. Furthermore, other maintenance concepts such as failure-based and age-/use-based maintenance or pure repair strategies without inspection can be handled as well.
Deterioration processes can be modelled as continuous stochastic processes. They are transformed to discrete processes as the decision process is modelled as a discrete process. This discretisation diminishes modelling and solution complexity and increases modelling potentiality.
With an example we have demonstrated that one can profit considerably from analysing different strategy concepts and optimising parameter values using a model such as the one described in this paper. In doing so, we have showed that the concept of condition-based inspection and repair strategies can be more profitable than other concepts. As a remark we could add, however, that their implementation usually makes more demands upon the maintenance planner than other concepts.
Model implementation and further theoretical development continue steadily; not all possibilities of the approach have been fully explored yet. Model extensions are being investigated. We have developed the OPTIMON software package which supports the approach and has been used by the Netherlands Department of Public Works for some time now.

6 Notation

N	number of condition levels
F	state of failure
i	condition level of a component
q_1	probability of a failure revealing itself without inspection
q_2	probability of a hidden failure revealing itself after all (per unit of time)
$G(t)$	condition at time t
$g(t)$	average condition at time t
b	uncertainty parameter of the condition at time t
C	cost of one inspection
$CR(i)$	cost of one repair (replacement) when the component is in condition level i
$CO(i)$	operating costs per unit of time in working condition i
$CO(i)$	$i = N, N + 1$; failure costs per unit of time
CF	additional failure costs per failure
TI	duration of one inspection
$TR(i)$	duration of one repair (replacement) if the component is in condition level i
$TD(i)$	time between the end of an inspection and the beginning of the subsequent repair if the component is found to be in condition level i
r	interest rate
α	discount rate
R	maintenance strategy of the control-limit type
π_0	condition level at which (or worse) repair is prescribed
π_i	length of inspection interval in units of time if the last known condition level is i
m	number of time units passed since the condition level has become known
a	action indicator
M_i	maximum length of inspection interval in units of time if the last known condition level is i
$\mathbf{S}$	set of states the decision process can be in
s or s'	a particular state
$P\{s \overset{(t)}{\rightarrow} s' \mid R\}$	probability of a transition from state s to state s' in t steps if in every state an action according to strategy R is taken
$P\{s \rightarrow s' \mid a_R\}$	probability of a transition from state s to state s' in one step if in state s an action according to strategy R is taken
$C_s(a_R)$	expected transition costs if in state s an action according to strategy R is taken
$T_s(a_R)$	expected transition time if in state s an action according to strategy R is taken
$v_s(R)$	expected total discounted costs if the process starts in state s and strategy R is adhered to

$T_R(s;a)$ expected total discounted costs if the process starts in state s and in this state s an alternative action is taken instead of the action prescribed by strategy R

7 References

BURGER, H.H., HONTELEZ, J.A.M., WIJNMALEN, D.J.D. (forthcoming in 1995), Optimum condition-based maintenance policies for deteriorating systems with partial information, *Journal of Reliability Engineering and Systems Safety*

HONTELEZ, J.A.M., WIJNMALEN, D.J.D. (1991), *Optimisation of maintenance costs of the components of a system* (in Dutch), TNO Physics and Electronics Laboratory, The Hague, Report FEL-89-C379

TIJMS, H.C. (1986), *Stochastic modelling and analysis: A computational approach*, Wiley, Chichester

TIJMS, H.C., DUYN SCHOUTEN, F.A. VAN DER (1984), A Markov decision algorithm for optimal inspections and revisions in a maintenance system with partial information, *European Journal of Operational Research*, vol. 21, pp. 245-253

WIJNMALEN, D.J.D. (1992), *Minimising discounted maintenance costs: OPTIMON with discount rate* (in Dutch), TNO Physics and Electronics Laboratory, The Hague, Report FEL-92-C250

WIJNMALEN, D.J.D., HONTELEZ, J.A.M. (1992), Review of a Markov decision algorithm for optimal inspections and revisions in a maintenance system with partial information, *European Journal of Operational Research*, vol. 62, pp. 96–104

WIJNMALEN, D.J.D., HONTELEZ, J.A.M. (1993), Opportunistic condition-based repair and inspection strategies for components of a multi-component maintenance system with discounts, manuscript submitted for publicaton in *European Journal of Operational Research*

8 Appendix

The formulas of the transition probabilities are as follows:

a. $P\{(i,m)\rightarrow(i,m+1)\,|\,0\}=\dfrac{1-r_{iN+1}(m+1)}{1-r_{iN+1}(m)}$ $\quad i=1,\ldots,N-1;\ m=0,\ldots,M_i-1$

b. $P\{(i,m)\rightarrow(N+1,0)\,|\,0\}=\dfrac{r_{iN+1}(m+1)-r_{iN+1}(m)}{1-r_{iN+1}(m)}$ $\quad i=1,\ldots,N-1;\ m=0,\ldots,M_i-1$

c. $P\{(i,m)\rightarrow(j,0)\,|\,1\}=\dfrac{r_{ij}(m+TI')}{1-r_{iN+1}(m)}$ $\quad i=1,\ldots,N-1;\ j=i,\ldots,N;\ m=1,\ldots,M_i$

d. $P\{(i,m)\rightarrow(i,N+1)\,|\,1\}=\dfrac{r_{iN+1}(m+TI')-r_{iN+1}(m)}{1-r_{iN+1}(m)}$ $\quad i=1,\ldots,N-1;\ m=1,\ldots,M_i$

e. $P\{(i,0)\rightarrow(1,0)\,|\,2\}=1$ $\quad i=2,\ldots,N+1$

Note that in deterioration probabilities $r_{ij}(m+TI')$, TI' is defined as follows:

$$TI'=\begin{cases}0 & \text{if deterioration does not continue during inspection}\\ TI & \text{if deterioration continues during inspection}\end{cases}$$

We shall give some examples of expected transition times and costs. For a complete set of formulas we refer to Wijnmalen (1992).

f. $C_{(i,0)}(0) = CO(i)$ $\qquad i = 1, \ldots, N-1$

g. $C_{(i,m)}(0) = \dfrac{\sum_{j=i}^{N} r_{ij}(m) \cdot CO(j)}{1 - r_{iN+1}(m)}$ $\qquad i = 1, \ldots, N-1;\ m = 1, \ldots, M_i$

h. $C_{(i,m)}(1) = CI$ $\qquad$ if $TI = 0$, $i = 1, \ldots, N-1;\ m = 1, \ldots, M_i$

i. $$C_{(i,0)}(2) = \begin{cases} CR(i) & \text{no delay; } i < N \\ CR(i) + CF & \text{no delay; } i = N, N+1 \\ \alpha^{TD(i)} \cdot CR(i) & \text{no deteriation during delay; } i < N \end{cases}$$

j. $T_{(i,m)}(0) = 1$ $\qquad i = 1, \ldots, N-1;\ m = 0, \ldots, M_i - 1$

k. $T_{(i,m)}(1) = TI$ $\qquad$ no deterioration during inspection

l. $T_{(i,0)}(2) = TD(i) + TR(i)$ $\qquad$ no deterioration during delay

The set of equations (2) is:

$$\begin{cases} v_{(i,m)} = C_{(i,m)}(0) + \alpha \cdot P_{(i,m)(i,m+1)}(0) \cdot v_{(i,m+1)} + \alpha \cdot P_{(i,m)(N+1,0)}(0) \cdot v_{(N+1,0)} & 1 \le i < \pi_0,\ 0 \le m < \pi(i) \\ v_{(i,0)} = C_{(i,0)}(2) + \alpha^{T_{(i,0)}(2)} \cdot v_{(i,0)} & \pi_0 \le i \le N+1 \\ v_{(i,\pi(i))} = C_{(i,\pi(i))}(1) + \alpha^{T_{(i,\pi(i))}(1)} \cdot \sum_{j=1}^{N+1} P_{(i,\pi(i))(j,0)}(1) \cdot v_{(j,0)} & 1 \le i < \pi_0 \end{cases}$$

After reduction, we obtain the set (3), where

$$A_i = \sum_{k=0}^{\pi(i)-1} \alpha^k \cdot (1 - r_{iN+1}(k)) \cdot C_{(i,k)}(0)$$

$$+ \alpha^{\pi(i)} \cdot (1 - r_{iN+1}(\pi(i))) \cdot C_{(i,\pi(i))}(1)$$

$$+ \sum_{k=1}^{\pi(i)-1} \alpha^{k+1} \cdot (r_{iN+1}(k+1) - r_{iN+1}(k)) \cdot C_{(N+1,0)}(2)$$

$$+ \alpha^{\pi(i) + T_{(i,\pi(i))}(1)} \cdot \sum_{j=\pi_0}^{N+1} r_{ij}(\pi(i) + TI') \cdot C_{(j,0)}(2)$$

$$B_i = \alpha^{T_{(N+1,0)}(2)} \cdot \sum_{k=1}^{\pi(i)-1} \alpha^{k+1} \cdot (r_{iN+1}(k+1) - r_{iN+1}(k))$$

$$+ \alpha^{\pi(i) + T_{(i,\pi(i))}(1)} \cdot \sum_{j=\pi_0}^{N+1} \alpha^{T_{(j,0)}(2)} \cdot r_{ij}(\pi(i) + TI')$$

$$C_{ij} = \alpha^{\pi(i) + T_{(i,\pi(i))}(1)} \cdot \sum_{j=i}^{\pi_0 - 1} r_{ij}(\pi(i) + TI')$$