This paper describes specific computational approach to reliability analysis, which is based on a state-transition model. By incorporating elements of both the deterioration process and the maintenance activities (inspections and repairs) in a semi-Markov model, a common computational foundation has been created for various dependability studies that can investigate different maintenance scenarios. Having available some basic model it is possible to adjust its parameters so that it represents some hypothetical new maintenance policy and then to examine an impact that the change to the new policy has on various reliability characteristics of the system. Particularly, this paper discusses an extension of the generic adjustment algorithm to specific situations of the so-called model saturation when, as a result of tweaking the model towards higher repair frequencies, sum of repairs probabilities in the states reach the maximum value and there is no room for further increase. The general idea is to modify the model in such cases by forcing some non-zero value of a repair probability in those states where it was zero initially but in a manner that would not destroy the overall model behaviour. This extension allows to successfully evaluating a class of cases that has not been properly handled by the generic method and broadens the range of dependability studies that can be effectively evaluated.

Keywords: state-transition deterioration model, semi-Markov process, model adaptation, maintenance analysis, model adjustment

1. Introduction

Cost-effective maintenance is the crucial point in management of any complex contemporary technical system. Selecting the optimal maintenance strategy is not an easy task and often requires extensive analysis of reliability, security, safety and economic aspects. Finding a reasonable balance between extensive and frequent maintenance actions on one side and redundant and excessive maintenance expenses on the other is the key point in reliable yet cost-effective system operation.

The subject of this paper is connected with original methodology that assists a person who decides about maintenance activities by evaluating risks and costs associated with choosing different maintenance strategies. Instead of searching for a globally optimal solution to a problem: “what maintenance strategy would lead to the best dependability parameters of system operation incurring the minimal cost”, in this approach different maintenance scenarios can be examined in “what-if” studies and their reliability and economic effects can be compared so that a person managing the maintenance is assisted in making informed decisions ([1–3]).

Our method of deterioration representation based on Markov models has been presented originally in [1] and its specific extensions are further described in [4–8]. Additionally, presentations in [9] and [10] concentrated on one important aspect of the methodology: fully automatic adjustment of the model to possible modifications of the maintenance policy that are often required in studies requested by the user. In this work we extend the research in this context with additional research related to the problem of the so-called model saturation that may occur when increased repair frequencies are requested and, furthermore, we propose an automatic modification mechanism that remedies the saturation limits.

The main content of the paper is divided in two parts. The first one (section 2) briefly summarizes the method of automatic model adjustment to the requested repair frequencies, which is the core task in maintenance studies, while the second part (section 3) discusses specific issues of the adjustment when probability saturation occurs and proposes an extension that circumvents these kinds of problems.

2. Modelling the Ageing Process

In this chapter we will briefly resume basic information about Markov models and the adjustment procedure that has been included in [9]. Further, in Chapter 3, this information is used for discussion of model saturation phenomenon and its proposed remedies.
2.1. Construction of the Model

Deterioration is a complex process and its modelling is not an easy task. In the literature there are numerous approaches proposed that, in general, try to encompass the three major factors that affect equipment wear: physical characteristics of the object under consideration, operating practices, and the maintenance policy. In the proposed solution especially the maintenance activities relate to the events and actions that should be properly incorporated in the representation and should be described with a distinctive set of parameters that would later be used in maintenance studies.

One of the approaches that can properly incorporate all the above suppositions about the aging process and maintenance activities is based on state-space (Markov) model ([12–17]). A Markov model consists of the states the equipment can assume in the process, and the possible transitions between them. The method described in this paper is based on a model of the Asset Maintenance Planner (AMP) that has been initially developed and implemented by George J. Anders and Henryk Maciejewski ([18–19]).

For structure of a typical AMP model see Figure 1. In the model, the deterioration progress is represented by a chain of deterioration states $D_1 \ldots D_K$, which then leads to the failure state $F$. In most situations, it is sufficient to represent deterioration by three stages: an initial ($D_1$), a minor ($D_2$), and a major ($D_3$) stage of deterioration ($K = 3$). This last is followed, in due time, by equipment failure ($F$). Other states are related to the maintenance activities: regular inspections ($I_s$ states) result in decisions to continue with, e.g., minor ($M_{s1}$) or major ($M_{s2}$) repair or to return to the deterioration state $D_s$ without any repair. The expected result of all repair actions is a single-step improvement in the deterioration chain.

\[
P_{sr} + \sum P_{s0} = 1.
\]

Figure 1. The state-transition model representing the deterioration chain with inspection and repair states (an example with two types of repairs is shown)

2.2. Adjusting the Model to Modified Repair Frequencies

Preparing the Markov model for some specific equipment is not an easy task and requires expert intervention. The goal is to create the model representing closely the real-life deterioration process known from the records that usually describe equipment operation under a regular maintenance policy with some specific frequencies of inspections and repairs. The model itself permits calculation of the repair frequencies and compliance of the computed and recorded frequencies is a very desirable feature that verifies trustworthiness of the model. In this sub-section, we will summarize the method of model adjustment proposed in [9] and [10] that aims at reaching such compliance. It can be used also for a different task of a fully automatic generation of a model for some hypothetical new maintenance policy with modified frequencies of repairs. Such a task is typical during evaluation of various maintenance scenarios.

In our analysis let the deterioration model under consideration consists of $K$ deterioration states and $R$ repairs. Also, let $P_{sr}$ denotes probability of selecting maintenance $r$ in state $s$ (assigned to the decision after inspection state $I_s$) and $P_{s0}$ represents probability of returning to state $D_s$ from inspection $I_s$ which corresponds to a situation when no maintenance is scheduled as a result of the inspection. The foremost condition that must be met at all times is that in all deterioration states $s = 1 \ldots K$:

\[
P_{s0} + \sum P_{sr} = 1.
\]
Let $F'$ represent the frequency of some repair $r$ as it is generated by the model. The problem of model adjustment can be formulated with various assumptions and with different goals in mind but in this approach it is defined as follows:

Given an initial semi-Markov model $M_0$, with internal structure representing deterioration, inspection and repair states as described above and producing the initial vector of repair frequencies $F_0 = [F_0^1, F_0^2 \ldots F_0^K]$, modify the probabilities $P_0^r$ assigned to transitions from inspections states $I$ so that the resulting model generates some requested vector of goal frequencies $F_G$.

There are various approaches that can be used in order to accomplish such model modification. In the proposed solution, a method of iterative approximations has been chosen in order to preserve an original construction of the model $M_0$ as much as possible. In this method a sequence of tuned models $M_0, M_1, M_2, \ldots M_N$ is evaluated with each consecutive model approximating desired goal with a better accuracy. Starting with $i = 0$ and the initial model $M_0$, the procedure consists in the following steps:

1° for the current model $M_i$ compute its vector of repair frequencies $F_i$

2° evaluate an error of $M_i$ as a distance between vectors $F_G$ and $F_i$

3° if the error is within the user-defined accuracy $\varepsilon$, consider $M_i$ as the final model and stop the procedure ($N = i$); otherwise proceed to the next step

4° create a new model $M_{i+1}$ by tweaking values of $P_i^r$, then correct $P_i^{r0}$ according to condition (1)

5° return to the step 1° for the next iteration.

Of all the steps 1–5, it is clear that adjusting probabilities $P_i^r$ in step 4° is the heart of the whole method. This is accomplished with the following two assumptions.

The first assumption is a restrictive condition: if the probability of some particular repair must be modified, it is modified proportionally in all deterioration states, so that during the adjustment the proportion between these repair probabilities over all states remains unchanged and is the same as in $M_0$:

$$\forall i, r \quad P_i^r : P_0^r : \ldots : P_0^R \approx P_i^r : P_{i-1}^r : \ldots : P_0^R.$$  

Now only $R$ scaling factors, denoted as $X_{i+1} = [X_{i+1}^1, X_{i+1}^2, \ldots X_{i+1}^R]$, must be found to compute all new probabilities and to create the next model $M_{i+1}$:

$$P_{i+1}^r = X_{i+1}^r \cdot P_i^r, \quad r = 1 \ldots R, \quad s = 1 \ldots K.$$  

Moreover, and this observation leads to the second assumption, although the frequency of a repair $r$ depends on the probabilities of all repairs (modifying probability of one repair changes, among others, state durations in the whole model; thus, it changes the frequency of all states) it can be assumed that, in a case of a single-step small adjustment, its dependence on repairs other than $r$ can be considered negligible and

$$F_i' = F_i(X_i', X_i^2 \ldots X_i^K) \approx F_i'(X_i')$$  

Now, generation of a tweaked model in step 4° is reduced to of solving a set of $R$ equations in the form:

$$F_i'(X_i') = F_G'$$  

and this can be accomplished with one of the standard root-finding numerical algorithms.

Additionally it should be pointed out that applying equation (3) with $X_{i+1} > 1$ may, in some state $s$, violate the condition

$$\sum_r P_{i+1}^r \leq 1.$$  

Such situation needs special tests and, if detected, execution of a scale-down transformation:

$$P_i^s = P_i^s / S_{Ds}, \quad S_{Ds} = \sum_{r=1}^{R} P_i^r.$$  

Efficient numerical methods for approximate solving the equation (5) and, consequently, implementing the whole adjustment procedure of the steps 1–5, were presented in [9] and [10].
3. Modifying the Model in Case of Probability Saturation

As it was discussed in [9] and [10], the above defined procedure can successfully adjust repair probabilities $P'$ when the goal (the $F_G$ vector) include the frequencies that are lower than the ones for which the model was created (the $F_0$ vector). That is, the method does not have problems if the model is transformed for studies like “what if some or all repairs are performed with lower frequency and deterioration rate is increased”. In these kinds of situations the $P'$ values need to be scaled down and the numerical algorithms are capable of their precise tuning so that the goal is reached in a reasonably limited number of iterations. On the other hand, adjusting the model to the repair frequencies that are substantially higher than the original ones may lead to so called model saturation – a condition in which repair probabilities reach the limit (6) and there is no room for further increase if the adjustment procedure is limited only to the simple scaling expressed by equation (3). In this situation bringing together two requirements such as tuning the model towards high repair frequencies, and at the same time keeping the modifications of the internal structure within a safe range, which does not break the proper relation with the original, it is a particular task that needs a devoted new development.

3.1. The Problem of Model Saturation

Discussion included in [6] investigated main challenges that are brought when the goal vector $F_G$ contains increased values of repair frequencies. The two main factors that were recognized were as follows: (1) although it may seem that in the initial (minor) deterioration state no repairs are performed after inspections, still some non-zero probabilities are required in $D1$ if purely hypothetical questions like “What if I start some repair twice as often as previously?” shall be allowed; (2) including an option of not doing any repair after inspection in the later deterioration states, albeit with small probability, is also desirable because it increases ability of the model to represent diverse maintenance configurations found in the studies.

For the purpose of this presentation, we will illustrate these issues with an example of two models with three deterioration states and two repairs: minor (index 1) and major (index 2). Thus, there are a total of 6 repair probabilities in the model ($P_{11}$ and $P_{12}$ in deterioration state $D1$, $P_{21}$ and $P_{22}$ in $D2$, $P_{31}$ and $P_{32}$ in $D3$) that will be fine-tuned by the procedure. Initial distribution of the probabilities is presented in Table 1. The model 1 has been created with assumption that although there are no repairs in the first state $D1$, when the equipment is in subsequent states $D2$ and $D3$ every inspection leads to some sort of repair and in these states the totals $S_{D2} = S_{D3} = 1$ ($P_{20} = P_{30} = 0$). Looking at the probability distribution in each state it can be seen that in the medium deterioration state $D2$ the minor repair evidently prevails ($P_{21} = 0.8$) while in the major deterioration state $D3$ the distribution is in favour of the major repair with only a little more balanced allocation of probabilities ($P_{32} = 0.7$). The model 2 is a sibling of 1 with one important difference; repair probabilities in $D2$ and $D3$ are lowered to, respectively, 80% and 95% of the values taken from 1, which means that after inspections in these states it is possible to return to $Ds$ without undertaking any repair: $P_{20} = 0.2$ and $P_{30} = 0.05$.

Table 1. Repair probabilities in the sample models used as examples in this work

<table>
<thead>
<tr>
<th>Deterioration state</th>
<th>$D1$</th>
<th>$D2$</th>
<th>$D3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repair probability</td>
<td>$P_{11}$</td>
<td>$P_{12}$</td>
<td>$P_{21}$</td>
</tr>
<tr>
<td>Model 1</td>
<td>0</td>
<td>0</td>
<td>0.80</td>
</tr>
<tr>
<td>Model 2</td>
<td>0</td>
<td>0</td>
<td>0.64</td>
</tr>
</tbody>
</table>

The following analyses will consist in generation of a sequence of models adjusted to goal frequencies created by modifying frequency of the major repair (index 2):

$$F_G = [F_0', \alpha \cdot F_0'],$$

where the factor $\alpha$ will increase from 0.5 (frequency of the major repair reduced by half) to 2.0 (the repair performed twice as often), in steps of 0.1. Values of $\alpha$ in the figures and in the following discussion will be expressed as %. Frequency of the major repair was selected as the varying parameter of $F_G$ just as an example, but equivalent results could be demonstrated with changing frequencies of the minor repair. The figures will include graphs presenting variations of repair probabilities $P''$ and their sums $S_{Ds}$ in deterioration states of the final adjusted models as functions of the $\alpha$ factor.
The problem of probability saturation is illustrated in Figure 2 which shows major repair probabilities in all states for models 1 and 2 tuned with the standard procedure described in the previous section (upper graphs) along with sums of probabilities over all states (lower graphs). Probability of the minor repair is not included to preserve space. Although it does not remain constant (see, for example, discussion in [6]) its variation does not demonstrate any significant aspects of saturation mechanism and just follow the general rules of repair inter-dependence.

The graphs show that both models can be successfully adjusted only up to the point of saturation which is reached for $\alpha = 100\%$ for the model 1 (i.e. the initial model is already saturated) and 120% for model 2 (in this particular case probabilities $P_{20} = 0.2$ and $P_{30} = 0.05$ leave room which is enough for 20% increase in frequency of the major repair). In both cases in the saturation points ($\alpha = 100\%$ and 120%) probabilities in states $D_2$ and $D_3$ sum up to unity and cannot be further increased, while in $D_1$ the $P_{12}$ is zero and applying the scaling factor as in equation (3) cannot produce any growth. On the other hand, the procedure has no problems with adjustment towards lower frequencies and in such cases the probabilities are scaled accordingly which only confirms discussions presented in [9] and [10].

3.2. Automatic Modification of the Model in Case of Saturation

The above example of unsuccessful tuning can be used also for illustration of the main idea of the proposed extension to the algorithm: if the model gets saturated during the adjustment iteration but there is still some state with null repair probability, the process can be continued in the same iterative way after some non-zero probability is introduced into this state. Such modification, though, goes far beyond the restrictive assumption expressed by equation (2) and, being a more serious invasion into the model structure, must be applied in a cautious and thoughtful manner.

As it was investigated in [7], the following two issues must be taken into account: (1) forcing non-zero probability in some state before it is absolutely necessary, i.e. prior to model saturation, instantly changes reaction to the adjustment iterations, hence may change the final result of the tuning also in cases when the standard procedure would be able to produce the correct result; (2) replacing the null value of $P^*$, even if delayed up to the moment of saturation, but with probability which is too high for the needs of the adjustment also may affect the final result in a way that is against the general idea of the conservative tuning which tries to preserve the structure of the original model with minimal possible modifications. Consequently, it is advantageous to minimize the value of the newly introduced probability even below...
the anticipated level: the standard adjustment procedure that continues afterwards would increase it accordingly on the cost of extra iterations anyway, whereas the overestimated value would adversely corrupt the adjustment procedure and harm the result – the goal frequencies would be reached but with unwanted changes of the model structure ([7]).

Taking all these subtle aspects into account, the following approach has been selected as the optimal and robust solution. In order to limit the changes of the already implemented method, the overall process of the 5 steps outlined in section 2.2 remains unchanged while only the internal procedure of probability tweaking in step 4° becomes extended. The procedure receives as the input the current $M_i$ model and produces as the output the next $M_{i+1}$ model in a sequence of operations that can be illustrated by the following pseudo-code:

1. procedure $TweakProbs(M)$
2. \hspace{1em} $EstimateTweakFactors(M, X)$; \hspace{1em} // ...with numerical method NOLA, secant or falsi
3. \hspace{1em} if not $ModelSaturated(M)$ then
4. \hspace{2em} // Model IS NOT saturated, use the standard procedure
5. \hspace{3em} for each $s$ in $M.DeteriorationSates$ do
6. \hspace{4em} for each $r$ in $M.Repairs$ do $P'' := P'' \cdot X_r$;
7. \hspace{4em} $S = \sum_r P''$;
8. \hspace{4em} if $S > 1$ then
9. \hspace{5em} for each $r$ in $M.Repairs$ do $P'' := P'' / S_{Di};$
10. \hspace{4em} end if;
11. \hspace{3em} next $s$;
12. \hspace{1em} else
13. \hspace{2em} // Model IS saturated
14. \hspace{3em} double $dAvgInc[R]$; \hspace{1em} // an array for predicted average increases of $P''$
15. \hspace{3em} $EstimateAverageProbIncrease(M, X, dAvgInc);$;
16. \hspace{3em} for each $s$ in $M.DeteriorationSates$ do
17. \hspace{4em} for each $r$ in $M.Repairs$ do
18. \hspace{5em} if $P'' > 0$ then
19. \hspace{6em} // The probability is above zero, apply the regular scaling
20. \hspace{6em} $P'' := P'' \cdot X_r$;
21. \hspace{6em} else
22. \hspace{7em} // The probability is zero, force the positive value
23. \hspace{7em} $P'' := dAvgInc[R]$;
24. \hspace{6em} end if;
25. \hspace{5em} next $r$;
26. \hspace{4em} $S = \sum_r P''$;
27. \hspace{4em} if $S > 1$ then
28. \hspace{5em} for each $r$ in $M.Repairs$ do $P'' := P'' / S_{Di};$
29. \hspace{5em} end if;
30. \hspace{4em} next $s$;
31. \hspace{3em} end if;
32. \hspace{1em} end procedure;

The operation begins as the previous unmodified method: initially, in the line no. 2 the scaling factors $X'$ are computed using any of the numerical approximation methods (NOLA, secant or falsi), as it was described in detailed discussions of [9] and [10]. Afterwards, the conditional instruction in the line no. 3 divides the rest of the code into two parts which are executed either for non-saturated (lines 4 ÷ 10)
or for saturated (lines 12 ÷ 26) models. The model is considered saturated (which is tested as the condition of the line 3) if in all states \( S_{D_1} = 1 \) (probabilities reached their maximum) or \( S_{D_2} = 0 \) (probabilities are zero and cannot be increased by simple multiplication) and any of the \( X \) factor is greater than 1 (if all probabilities are to be decreased during the adjustment the saturation is not an obstacle and the standard method should not be altered). If the model is not saturated, the standard procedure just multiplies the probabilities by the tweaking factors (second part of the line 5 includes the equation (3)) and, per every state, the new values are scaled down if their sum exceeds the limit (lines 6 ÷ 9 implement the equation (7)).

Actual extension to the algorithm is included in the special processing path of the lines 12 ÷ 26 which is executed if the model has reached the saturation state. In this case, in the beginning values of predicted average increases of the probabilities are computed in line 13 and stored in auxiliary \( dAvgInc[\cdot] \) array. Details of this computation has been omitted for brevity, but the code actually repeats the standard path similar to that of lines 4 ÷ 10 with the only difference that the new values are not assigned but they are used for storing the \( P_{sr} \) changes in elements of the \( dAvgInc[\cdot] \) array. If, afterwards, for any repair \( r \) the accumulated change is negative (a common case when this particular repair frequency should be lowered) it is replaced with zero while positive values are later used in forcing non-zero \( P_{sr} \) values. The rest of the processing is executed on state-by-state basis and is expressed with a single for...each instruction spanning lines 14 ÷ 26. In every state, each probability is either multiplied by the scaling factor if it is positive and such multiplication does lead to any increase (line 17 analogous to the equation (3)), or it is replaced with its pre-computed average increase in other states \( dAvgInc[r] \). The line no. 19 implements the actual operation of replacing a zero value of \( P_{sr} \) with a positive one, which is the essence of the method. Afterwards, in lines 22 ÷ 25 the probabilities are scaled down if their sum exceeds the limit, although this conditional operation usually will not be executed in states where the probability increase was done as the condition in the line no. 23 will not be fulfilled.

Usually, once a non-zero probability has been forced into the model, the further iterations operate on a non-saturated model and they can proceed only with the regular scaling operations of the lines 4 ÷ 10.

4. Conclusions

In this work we were dealing with a deterioration modelling methodology which was designed to help in choosing effective yet cost-efficient maintenance policy. Incorporating in the state-transition semi-Markov model elements of both deterioration characteristics of the object and the maintenance activities (inspections and repairs), a common computational foundation has been created for various dependability studies on possible maintenance scenarios. Specifically, this paper presented an extension of the model adjustment algorithm which was proposed and refined in our previous works. The general idea is to modify the model during the iteration by forcing a value greater than zero for a repair probability in situation when this probability reach the limit in other states, i.e. the model saturates. This extension allows to successfully evaluating a class of cases that was not properly handled by the original method and broadens the range of dependability studies that can be effectively evaluated.

The proposed approach strives to be as conservative as possible with regard to the amount of alterations introduced to the existing model. While the original method constrains the adjustment operations so that the distribution of the repair probabilities over all deterioration states is altered to the least possible degree, the modification introduced by this extension is a far more significant one and must be applied in a very cautious manner in order to avoid deformation of the model and corruption of the produced results. Consequently, there is a growing need for methods that would evaluate trustworthiness of the generated results, for example definition of new metrics of the model that would quantitatively assess extensions of its modification and would allow estimating the range of its valid use.

References


