PK8207 - Lecture memo

Jørn Vatn
Email: jorn.vatn@ntnu.no

Updated 2020-04-27

Grouping and opportunity maintenance

Introduction

In classical maintenance optimization the objective is to find the optimum frequency of maintenance of one component at a time. However, in the multi-component situation there exist dependencies between the components, e.g., they may share a common set-up costs (economy of scope), the costs may be reduced if the contract to a maintenance contractor is huge (economy of scale), etc.

In this presentation we will introduce some rather simple approaches for maintenance grouping and opportunity maintenance. We basically consider the following cost elements:

- Man-hour costs and material costs related to preventive maintenance of each component
- Set-up costs to get access to the components to be maintained, and by paying the set-up costs access to several components is obtained. We limit the scope to consider a one level structure of set-up costs, meaning that the set-up cost is the same for all components. In a multi level structure the set-up cost could be split into a general set-up cost for accessing e.g., a location/cite, and further into set-up cost for a group of components related to e.g., preparing of the work for these components.
- Costs of taking a component out of service. These costs are included in the set-up costs from a modelling point of view.
- Man-hour costs and material costs related to corrective maintenance. Typically set-up costs can not be shared by other components unless preventive maintenance is advanced (opportunity maintenance).
- Costs related to the effect of a failure, i.e., punctuality, unavailability, safety and material damage costs

We often distinguish between the static and the dynamic planning regimes. In the static regime the grouping is fixed during the entire system lifetime, whereas in the dynamic regime the groups are re-established over and over again. The static grouping situation may be easier to implement than the dynamic, and the maintenance effort is constant, or at least predictable. The advantage of the dynamic grouping is that new information, unforeseen events, etc., may require a new grouping and changing of plans.

The presentation here discusses how we can formalize the optimization of maintenance grouping. I.e., we seek to group maintenance tasks so that total costs are minimized. To summarize the difference between static and dynamic grouping we have:

- Static grouping where the groups are fixed
  - It is always the same maintenance task that are included in the same group, and each maintenance group is performed at a fixed interval.
  - In the work order system one group is specified as one work order repeating every τ_i unit of time.
- Dynamic grouping where we create the groups "on the fly"
  - The time of next maintenance is recalculated in principle continuously
  - The set of maintenance tasks going into a group is varying form time to time
  - In principle we can plan for several groups ahead, but often we only consider the first group of tasks
  - We can update the plan if we get new information, or there are additional opportunities to carry out maintenance
  - The downside is significantly more administrative work and challenges in relation to staff planning.

For an introduction to maintenance grouping we refer to Wildeman (1996) who discusses these different regimes in detail.

Maintenance tasks are here preventive tasks where the base interval is calendar controlled or controlled by runtime. At the end of the presentation we also discuss condition-based maintenance.

The costs of disassembling and re-assembling are here included in the set-up cost. In the model presented we also assume that the set-up costs are the same for all activities. It is further assumed that there is one and only one maintenance activity related to each component. This simplifies notation because we then may alternate between failure of component i and executing maintenance activity i where there is a unique relation between component and activity. The basic notation to be used is below. The terms maintenance task and maintenance activity are used interchangeably. Table 1 shows the notation used. Note that t is used to represent calendar/global time or accumulated mileage for a car or a train. x is used to represent local time, i.e., time since last maintenance.

Static grouping

For static grouping, we distinguish between indirect and direct grouping:

- Indirect grouping means that the groups are not established by a direct rule, but that the groups are established based on a principle. This principle is that an activity can be executed on each maintenance opportunity, every second opportunity and so on. How often to be executed is then the optimization challenge.
- Direct grouping means that the groups are established by investigating the intervals one by one and form groups of activities activities having approximately the same interval.

Indirect static grouping

The indirect grouping principle is that the time of each activity is determined indirectly by specifying how often the task is performed relative to a fixed repetitive time of maintenance. The situation now is as follows:

- There is a possibility to do preventive maintenance at point of times $T, 2T, 3T, \ldots$
- For component i this opportunity is utilized every l_i'th time, i.e., the interval between maintenance for this component is $l_i T$
- The challenge is to determine T and $l_i, i = 1, 2, \ldots$

For a given value of T and $l_1, l_2, \ldots$ the expected cost per unit time is:

$$\begin{aligned} C(T, l_1, l_2, \ldots) &= S/T + \sum_{i=1}^{n} \left[c_i^{\mathrm{P}} + M_i(l_i T) \right] / (l_i T) \\ &= S/T + \sum_{i=1}^{n} \left[c_i^{\mathrm{P}} / (l_i T) + c_i^{\mathrm{U}} \lambda_{\mathrm{E},i}(l_i T) \right] \end{aligned} \tag{1}$$

where $M_i(x)$ is the total failure related cost in a period of length x since last maintenance.

Table 1: Notation

c_i^{P}	Planned maintenance cost, exclusive set-up cost for activity i. Typically the costs of replacing one unit periodically
c_i^{U}	Unplanned costs upon a failure of component i. These costs include the corrective maintenance costs, safety costs, punctuality costs, unavailability costs and costs due to material damage.
S	Set-up cost, i.e., costs for preparations, access etc which can be "shared" by several PM activities
$\lambda_{\mathrm{E},i}(x)$	Effective failure rate for component i. Here the argument x represents local time since last maintenance
$M_i(x)$	$= xc_i^{\mathrm{U}}\lambda_{\mathrm{E},i}(x)$ = Accumulated expected costs due to failures in a period $[0,x)$ for component i maintained at time 0, exclusive planned maintenance cost
$\Phi_i(x,k)$	$= [c_i^{\mathrm{P}} + S/k + M_i(x)]/x$ = Expected total costs per unit of time for component i for a maintenance cycle of length x if setup costs are shared by k activities
$x_{i,k}^*$	Maintenance interval that minimizes $\Phi_i(x,k)$ if setup costs are shared by k activities
$\Phi_{i,k}^*$	Minimum cost for a component i maintained at optimal interval
k_i	Average number of components sharing the set-up costs for the i'th component, i.e., the i-th component is in average maintained together with $k_i - 1$ other components
Φ_i^*	Average minimum costs per unit time over all k-values
x_i^*	Optimum value of x_i over all k-values. x_i^* is measured since last maintenance on component i
t_0	Point of time when we are planning the next group of activities. Initially $t_0 = 0$. t_0 is measured in running time since $t = 0$.
x_i	Age of component i at time t_0, i.e., time since last preventive maintenance activity
t_i^*	$t_{i,}^* = t_0 + x_i^* - x_i$ = optimum time for execution in the average situation
$G(g)$	Candidate group, i.e., the set of the first g components to be maintained according to individual schedule with $t_{i,\mathrm{Av}}^*$ as the basis for due time
l_i	How often a component is utilizing the maintenance opportunity in static indirect grouping
N	Number of activities/components
T	For dynamic grouping T is the end of planning horizon, i.e., we are planning from $t_0 = 0$ to T. For indirect static grouping T is used as the lowest interval.
T_j	Interval for group j in static direct grouping.

Heuristic for indirect static grouping

Minimizing Equation (1) is a mixed-integer optimization problem. Generally such problems need to be solved by heuristics when N becomes large. The following heuristic is suggested to find a reasonably good solution:

1. For each activity i we find the value of τ_i which minimizes $C(\tau_i) = \left(S + c_i^{\mathrm{P}}\right)/\tau_i + c_i^{\mathrm{U}} \lambda_{\mathrm{E},i}(\tau_i)$
2. An initial value of T is set equal to the lowest value of the τ_i-values
3. Chose $l_i \approx \tau_i/T$ (nearest integer)
4. Keep the l_i's fixed, and minimize $C(T, l_1, l_2, \ldots)$ with respect to T
5. GoTo 3 and change the l_i's ± 1 one by one to search for better solutions
6. An approximate optimal solution is found when the iteration scheme does not improve the solution, i.e., we do not find a solution with a lower expected cost

Exercise 1

Consider a situation where we have 4 components. We will establish a standard indirect static grouping strategy.

The following data is provided: $S = 2, c_1^P = 2, c_2^P = 1, c_3^P = 3, c_4^P = 1. c_1^U = 5, c_2^U = 50, c_3^U = s_4^U = 10. \mathrm{MTTF}_1 = 4, \mathrm{MTTF}_2 = 3, \mathrm{MTTF}_3 = 3, \mathrm{MTTF}_4 = 5, \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 3$.

Find tentative optimal intervals for each component if they are maintained individually and where we assume that we do not have to pay the set-up cost. Use this to find tentative values for l_i and T. Then try some iterations to see if a better solution can be found. Note, in the heuristic we proposed to find the individual solutions assuming we pay the entire set-up cost. In this exercise a slightly different approach was proposed, i.e., that we do not have to pay the set-up cost. The two approaches should converge to the same result, i.e.,

Exercise 2

Consider the situation in Exercise 1. Repeat the analysis if you in the initiation assume that we have to pay the set-up cost.

Direct static grouping

By direct grouping, the groups are selected directly by inspecting individual intervals. Tasks are now split into m non-overlapping groups, $G_1, G_2, \ldots$.

Activities in a group are maintained at the same time. The groups are established so that activities in a group have approximately equal intervals. For group j, we let T_j denote the interval for this group. Total expected costs per unit time is given by

$$C(T_1,T_2,T_3,\ldots,T_m) = \sum_{j=1}^{m}\left(S/T_j + \sum_{i\in G_j}\left[c_i^{\mathrm{P}}/T_j + c_i^{\mathrm{U}}\lambda_{\mathrm{E},i}(T_j)\right]\right) \tag{2}$$

Heuristic for direct static grouping

The following heuristic is proposed for obtaining a reasonable good solution:

1. For each activity find the value τ_i which minimizes $C(\tau_i) = (S + c_i^{\mathrm{P}})/\tau_i + c_i^{\mathrm{U}}\lambda_{\mathrm{E},i}(\tau_i)$
2. Sort in increasing order, i.e., $\tau_{(1)} \le \tau_{(2)} \le \ldots$
3. Look for natural clusters in the intervals, and let these forms groups $G_1, G_2, \ldots$
4. Given a split into groups, i.e., $G_j, j = 1,2,\ldots,m$, minimize the cost Equation (2) with respect to $T_1, T_2, \ldots, T_m$
5. GoTo 3 and varying the groups to look for better solutions, for example moving one activity from one group to another group, merge two groups, or split one group into two groups
6. An approximate optimal solution is found when the iteration scheme does not improve the solution.

Exercise 3

Consider the situation in Exercise 1. Propose

Dynamic grouping

In dynamic grouping there are no fixed group. At a given point of time, t_0, we start forming the next group based on "individual" due dates. Figure 1 illustrates the situation for four components maintained at t_1, t_2, t_3 and t_4 in the past, where the due dates t_1^*, t_2^*, t_3^* and t_4^* are based on the individual optimal intervals x_1^*, x_2^*, x_3^* and x_4^*.

In the optimization we consider the time from now on, t_0, up to the end of planning horizon, T. Given the information we have at time t_0 we can form the next group and when to execute the corresponding maintenance activities. We can also form the second group, the third group etc. But then

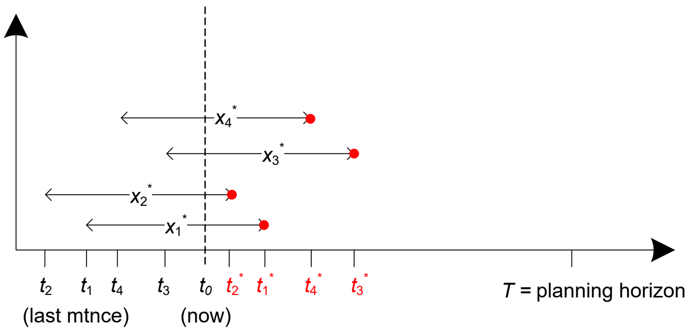

Figure 1: Grouping - Due dates. The t_i's are the point of time of last maintenance, the x_i^*'s are the individual optimal intervals, and the t_i^*'s are the corresponding due dates.

we should realize that we might get new information later on, and hence have to reschedule some future groups.

In this situation there is no single cost equation to optimize. We will structure the cost elements and then propose a heuristic for forming groups.

For each component i there is an expected time dependent cost which is a function of the time since the last preventive maintenance activity, i.e., $M_i(x)$. In order to establish $M_i(x)$ we need (i) to establish the accumulated expected number of failures in the period $[0, x)$, (ii) we need to specify the expected corrective maintenance costs for the repair of each failure, and (iii) we have to specify the impact of the failure on safety, production, etc., and quantify these into cost figures. In the model presented here we assume that the effective failure rate, $\lambda_{\mathrm{E},i}(x)$ may be established for the different failure characteristic, and maintenance strategies (e.g., periodic replacement and condition monitoring). Next the costs associated with a failure of component i may in principle be found by risk modelling, reliability modelling. The result of such modelling is one figure for the expected unplanned cost of failure, i.e., c_i^{U}. We have

$$M_i(x) = x c_i^{\mathrm{U}} \lambda_{\mathrm{E},i}(x) \tag{3}$$

The planned costs comprise the costs of executing the maintenance on component i (c_i^{P}) and set-up costs (S) of getting access to the component. The set-up costs may in general be shared with $k-1$ other activities. The average contribution to the total costs for component i per unit time is given by:

$$\Phi_i(x,k) = \left[c_i^{\mathrm{P}} + S/k + M_i(x)\right]/x \tag{4}$$

If the grouping was fixed, i.e. static grouping, the optimization problem would just be to minimize $\Phi_i(x,k)$ wrt x for all k components maintained at the same time. For dynamic grouping the mathematical challenge is now to establish the grouping either in a finite or infinite time horizon. In addition

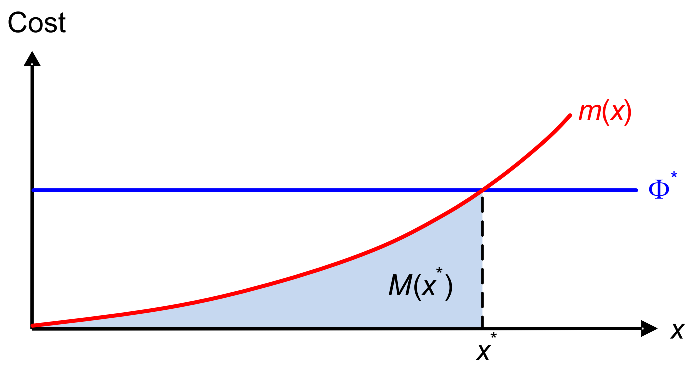

Figure 2: Marginal cost consideration

to the grouping, we also have to schedule the execution time for each group (maintenance package). The grouping and the scheduling can not be done separately. Generally, such optimization problems are NP hard (see Garey and Johnson, 1977, for a definition), and heuristics are required. Before we propose our heuristic we present some motivating results.

Let $\Phi^*_{i,k}$ be the minimum average costs when one component is considered individually, and let $x^*_{i,k}$ be the corresponding optimum x value. It is rather easy to prove that

$$m_i(x^*_{i,k}) = M'_i(x^*_{i,k}) = \Phi^*_{i,k} \tag{5}$$

meaning that when the instantaneous expected unplanned costs per unit time, $m_i(x)$, exceeds the average costs per unit time, maintenance should be carried out. The way to use the result is now the following. Assume we are going to determine the first point of time to execute the maintenance, i.e., to find t^*_{i,k_i} starting at $t = 0$. Further, assume that we know the average costs per unit time, Φ^*_{i,k_i} but that we have for some reason "lost" or "forgotten" the value of the optimal interval, x^*_{i,k_i}. What we then can do is to find t such that $m_i(t) = M'_i(t) = \Phi^*_{i,k_i}$ yielding the first point of time for maintenance, see Figure 2 for an illustration. Then from time t and the remaining planning horizon we can pay $\Phi^*_{i,k}$ as the minimum average costs per unit time. This is the traditional *marginal costs* approach to the problem, and brings the same result as minimizing Equation (4).

The advantage of the marginal thinking is that we now are able to cope with the dynamic grouping. Assume that the time now is t_0, and x_i is the age (time since last maintenance) for component i in the group we are considering for the next execution of maintenance. Further assume that the planning horizon is $[t_0, T)$. The problem now is to determine the point of time $t(\geq t_0)$ when the next maintenance is to be executed. The total costs of executing the maintenance activities in a group is

$$C_{\mathrm{P}} = S + \sum_i c_i^{\mathrm{P}} \tag{6}$$

which we pay at time t. Further, the expected unplanned costs in the period $[t_0, T)$ is

$$C_{\mathrm{U}} = \sum_i [M_i(t - t_0 + x_i) - M_i(x_i)] \tag{7}$$

where x_i is the (local) age of component i at time t_0. Note that $M_i(t - t_0 + x_i)$ is the expected cost from the last maintenance of component i until it will be preventively maintained at time t. From this value we subtract the expected cost $M_i(x_i)$ already "paid" at time t_0. Note that at time t_0 we know the "history" of component i since the last maintenance, i.e., x_i time units ago. We might use this information to get a more correct expression for the expected cost in the interval $[t_0, t)$. It is not always easy to obtain such an expression, hence we often approximate with Equation (7).

For the remaining time of the planning horizon the total costs are

$$C_\infty = (T - t) \sum_i \Phi^*_{i,k_i} \tag{8}$$

provided that each component i can be maintained at "perfect match" with $k_i - 1$ activities the rest of the period. Since $\Phi^*_{i,k}$ depends on how many components that share the set-up cost, which we do not know at this time, we use some average value Φ^*_i. We assume that we know this average value at the first planning. To determine the point of time for maintaining a given group of components, say $G(g)$ with the g first activities we thus minimize:

$$c_1(t; g) = S + \sum_{i \in G(g)} \left[c^{\mathrm{P}}_i + M_i(t - t_0 + x_i) - M_i(x_i) + (T - t)\Phi^*_i \right] \tag{9}$$

The costs in Equation (9) depend on which components to include in the group of activities to be executed next. The more activities we include, the higher the costs will be. For some activities it might thus be cheaper to include them in groups to be executed later. For activities we do not include in this first group we assume that they will be maintained at their "optimum" time $t^*_i, > t$. The total contribution to the costs related to these activities in $[t_0, T)$ is:

$$c_2(t; g) = \sum_{i \notin G(g)} \left[c^{\mathrm{P}}_i + S/k_i + M_i(x^*_i) - M_i(x_i) + (T - t^*_i)\Phi^*_i \right] \tag{10}$$

provided they can be maintained at "perfect match" with other activities, i.e., the set-up costs are shared with $k_i - 1$ activities, and executed at time t^*_i. The total optimization problem related to the next group of activities is therefore to minimize:

$$\begin{aligned} c(t; g) = S &+ \sum_{i \in G(g)} \left[c^{\mathrm{P}}_i + M_i(t - t_0 + x_i) - M_i(x_i) + (T - t)\Phi^*_i \right] \\ &+ \sum_{i \notin G(g)} \left[c^{\mathrm{P}}_i + S/k_i + M_i(x^*_i) - M_i(x_i) + (T - t^*_i)\Phi^*_i \right] \end{aligned} \tag{11}$$

The idea is simple, we first determine the best group to execute next, and the best time to execute it. Further we assume that subsequent activities can be executed at their individual optimum. It is expected to do better by taking the second grouping into account when planning the first group, and not only treat the activities individually. See e.g., Buday et al (2005) for more advanced heuristics in similar situations to those presented here. The heuristic is as follows:

Step 0 - Initialization

This means to find initial estimates of k_i and use these k-values as basis for minimization of Equation (11). This will give initial estimates for x_i^* and corresponding and Φ_i^*. Finally the time horizon for the scheduling is specified, i.e., we set $t_0 = 0$ and choose an appropriate end of the planning horizon (T).

Step 1 - Prepare for defining the group of activities to execute next

Calculate tentative due dates $t_i^* = x_i^* + t_0 - x_i$ for all activities, and sort in increasing order. See Figure 1 for an illustration.

Step 2 – Establish the candidate groups

For $g = 1,2,\ldots,N$ we use the ordered t_i^*'s to find a candidate group $G(g)$ of size g to be executed next. If $t_g^* > \min_{i<g}(t_i^* + x_i^*)$ this means that at least one activity in the candidate group needs to be executed twice before activity g is scheduled which does not make sense. Hence, in this situation the last candidate group, $G(g)$ is dropped and we are not searching for more candidate groups at the time being.

Step 3 - Find optimum execution time for each candidate group

For each candidate group $G(g), g = 1,2,\ldots,$ minimize $c(t,g)$ in Equation (11) with respect to execution time t. Next choose the candidate group $G(g)$ that gives the minimum cost. This group should then be executed at the corresponding optimum time t.

Step 4 – Prepare for subsequent groups

We assume that all activities in the chosen candidate group are executed at time t. This corresponds to setting $x_i = 0$ for $i \in G(g), x_i = x_i + t - t_0$ for $i \notin G(g)$ and then update the current time, i.e., $t_0 = t$. If $t_0 < T$ GoTo Step 1, else we are done.

There are several ways to improve the algorithm. One intuitive improvement is to improve the estimates of k_i and corresponding x_i^* and Φ_i^* to be specified in Step 0. This is easy, since we in Step 4 get a new value of k for those activities included in the candidate group, and when the algorithm terminates we simply set k_i as the average for each activity i in the period $[0,T)$. We may then start over again at Step 0 with these new values of k_i.

Opportunity based maintenance

The dynamic scheduling regime presented above is a good basis for opportunity based maintenance. The scheduling we have proposed may be used to set up an explicit maintenance plan for the time horizon $[0,T)$. But even though the plan exists, we may consider changing it as new information becomes available, either in terms of new reliability parameter estimates, or if unforeseen failures occur. In operation, for any time t_0 we may update the scheduling of preventive maintenance.

Now assume that the due date t for the next scheduled maintenance of group $G(g)$ is larger than the current time t_0. Assume that a failure has occurred or there is another event occurring at time t_0 giving an opportunity to save the setup-cost S if we execute some preventive maintenance activities. Some, or all of the activities in $G(g)$ should now be considered for execution given the opportunity at time t_0. If $t_i^* \le t_0, i \le g$ this means that these activities have individual due dates in the past, hence it is obvious that these activities should be executed at this given opportunity.

Activities not scheduled in $G(g)$ should not be executed since they were not even included in a group to be executed later than t_0. The basic question is thus which of the remaining activities in $G(g)$ should be executed. Assume that we have found that it is favourable to execute the first $i-1 < g$ activities on this opportunity. The procedure to test whether or not activity i also should be executed is as follows:

- First we assume that all activities up to i are executed on this opportunity , i.e., we set $x_j = 0, j \le i$, and for activities above activity i, i.e., $j > i$, we set $x_j = t_0 - t_j$, and will evaluate the next group to be executed at some time $t' \ge t$:

- Let $C_1 = c_i^{\mathrm{P}} + \min_{t',g'} c(t',g')$, i.e., the best we can do if we decide to execute activity i at time t_0

- Next, we assume that only activities up to $i-1$ are executed at time t_0, i.e., $x_j = 0, j \le i-1$, and $x_j = t_0 - t_j, j \ge i$, where we also evaluate the next group:

- Let $C_2 = \min_{t',g'} c(t',g')$, i.e., the best we can do if we decide not to execute activity i at time t_0

- If $C_1 > C_2$ is it not beneficial to do activity i.

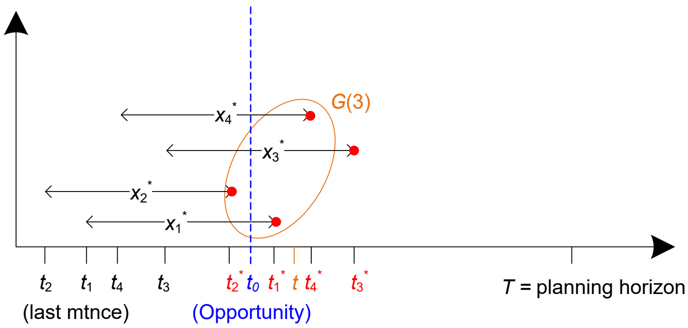

Figure 3: Opportunity maintenance. There is an opportunity at time t_0, where group $G(3)$ was scheduled for execution at time t.

If it was beneficial to do activity i at t_0 we should test for $i = i+1$ and repeat as long as $i \leq g$.

Figure 3 illustrates the situation. Assume that at the time of the opportunity we had already scheduled $G(3)$ was scheduled for execution at time t. The individual due date for activity 2 has been passed, hence activity 2 will be executed. The we consider if it pays of to execute also activity 1. If so, typically activity 3 and 4 will be a new group, say $G(2)$ to be executed at some time $t' > t$.

1 Synchronization of maintenance and production

We introduce a simple example to motivate for synchronization of maintenance and production. In a production line a packing machine is critical for the production. To simplify we only consider one activity, but the ideas could be combined with the theory of grouping discussed above.

In the example the mean time to failure is estimated to MTTF = 600 hours (≈ one month) if the machine is not preventively maintained. Failure times are assumed to be Weibull distributed and the ageing parameter is estimated to $\alpha = 4$. The cost of a preventive maintenance activity is $c_{\mathrm{PM}} = 5\,000$ NOKs. If the machine fails the total cost of corrective maintenance and lost production is $c_{\mathrm{U}} = 35\,000$. In the example we assume 24/7, i.e., continuous production using several shifts. The effective failure rate is approximated by: $\lambda_{\mathrm{E}}(\tau) = \left(\frac{\Gamma(1+1/\alpha)}{\mathrm{MTTF}}\right)^{\alpha} \tau^{\alpha-1}$, the cost function to minimize is thus:

$$C(\tau) = c_{\mathrm{PM}}/\tau + \lambda_{\mathrm{E}}(\tau)c_{\mathrm{U}} \approx c_{\mathrm{PM}}/\tau + \left(\frac{\Gamma(1+1/\alpha)}{\mathrm{MTTF}}\right)^{\alpha} \tau^{\alpha-1} c_{\mathrm{U}} \tag{12}$$

The optimal interval for predictive maintenance is found to be:

$$\tau^* = \frac{\mathrm{MTTF}}{\Gamma(1+1/\alpha)} \left(\frac{c_{\mathrm{PM}}}{c_{\mathrm{U}}(\alpha-1)}\right)^{1/\alpha} \approx \frac{600}{0.906} \left(\frac{5000}{35000 \times 3}\right)^{1/4} \approx 309 \text{ hours} \tag{13}$$

The optimal interval is found to be 309 hours which is 12.9 days. The production manager is not very happy with closing down production in order to perform preventive maintenance of the machine. However, every week there is a small production stop required for reconfiguration of the production setup to account for variants in the production. The production manager therefore proposes to utilize every second of these stops for carrying out the maintenance. Figure 4 shows the cost as a function of the maintenance interval, and it is rather obvious that shifting the interval from 309 hours to 336 (14 days) is not making much difference. In fact the yearly cost increases from 189 000 per year to 191 000 per year which is only an increase of 1%. The maintenance manager is not very happy since he did all the calculations and claims that 2 000 NOKs here and 2 000 NOKs there makes money. When the production manager puts the argument that the PM cost in fact would increase with at least 1 000 NOK's each time due to the interference with production, it is quite obvious that synchronization pays off.

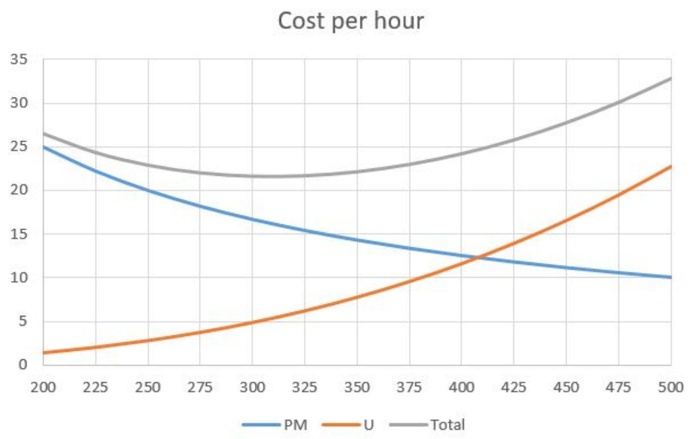

Figure 4: Cost per hour split into PM and unplanned (U) maintenance cost

In an intense production period the production manager is reluctant to carry out the scheduled maintenance that takes place every fourteen days. The machine has been operated for t = 14 days = 336 hours and it is due time for maintenance. But, rather than executing the PM activity on the due date, the production manager proposes to wait another week, i.e., to synchronize with the weekly production stop at the next occurrence. The maintenance manager hesitate to this proposal. The question is what are the arguments?

Since the challenge now is a “once in a life time" situation, the long term consideration does not apply. The approach is therefore to analyse the increase in maintenance related cost for this shift in maintenance and compare to the original strategy.

It will be sufficient to compare the first week, since for both the original strategy and the situation with the proposed postponing of maintenance from next week on the average cost is given by $C(\tau^*) = c_{\text{PM}}/\tau^* + \lambda_{\text{E}}(\tau^*)c_{\text{U}} = \Phi^* \approx$ 3 660 NOKs per week. For the original strategy the average cost for the coming week is also 3 660 NOK. When considering deferred maintenance we are facing a higher risk of failure since we are “climbing" on the increasing failure rate function. We will now assess the failure related cost from now on until the next PM. More generally we assume that t time units has elapsed since last preventive maintenance, and that we are considering to run the system for another x time units before the next PM activity is performed. The total unplanned failure cost in this entire time period $[0, t+x)$ is:

$$C_U(0, t+x) = \lambda_{\text{E}}(t+x) \cdot c_{\text{U}} \cdot (t+x) = M(t+x) \tag{14}$$

but we have already “paid" unplanned cost up to time t equal to $C_U(0,t) = \lambda_{\text{E}}(t) \cdot c_{\text{U}} \cdot (t)$, hence the cost in the coming x time units will be:

$$\begin{aligned} C_U(t, t+x) &= C_U(0, t+x) - C_U(0,t) = M(t+x) - M(x) \\ &= \lambda_{\text{E}}(t+x) \cdot c_{\text{U}} \cdot (t+x) - \lambda_{\text{E}}(t) \cdot c_{\text{U}} \cdot t \end{aligned} \tag{15}$$

In the example $t = 2$ weeks = 336 hours and $x = 1$ week = 168 hours yielding

$$\begin{aligned} C_U(t = 336, t+x = 336+168) &= C_U(0, 504) - C_U(0, 336) \\ &= \lambda_{\text{E}}(504) \cdot c_{\text{U}} \cdot (504) - \lambda_{\text{E}}(336) \cdot c_{\text{U}} \cdot (336) \\ &\approx 11\ 760 - 2\ 320 = 9\ 438 \text{ NOKs} \end{aligned} \tag{16}$$

This cost may be treated as a cost of a “high risk strategy" in the meaning that we are postponing a maintenance activity that in the optimal case should be executed at time t but we continue to “climb" on the increasing failure rate function. The cost of 9 438 NOKs should then be compared with the base case of $\Phi^* = 3$ 660 NOK. The increase in cost is almost 6 000 NOKs and this amount should then be compared to the value of not interrupting the production with maintenance.

Note that when calculating the unplanned failure cost in the period of deferred maintenance we have assumed that $c_{\text{U}} = 35\ 000$. This figure is an average cost of a failure including production interruption. In this case with intense production it is reasonable to assume that c_{U} could be even higher, causing the deferred maintenance strategy even more risky. The production manager therefore has to argue that the "production value" of deferred maintenance is at least 6 000 NOKs.

Predictive maintenance

The idea of predictive maintenance is to utilize the condition of a component and the future expected loads in order to judge the correct time for “hard" maintenance such as overhaul, replacement of worn parts, calibration and so on. Sensor technology is usually used to capture the condition of components or a system, and the term 'condition monitoring' is often used to describe the collection and analysis of state data relevant for predictive maintenance. It should be noted that manual inspection and use of "human sensors" to capture noise, smell, vibration could also be treated as condition monitoring. Concepts like digital twin and cyber physical systems are used to describe the situation where computerized system models interact with the physical systems in real time often by means of internet of things (IoT) and internet of services (IoS).

Cyber-physical systems (CPS) refers to smart systems that include engineered interacting networks of physical and computational components. The *term digital twin* refers to a digital replica of physical assets, processes and systems that can be used in real time for control and decision purposes. The digital twin representation is seen as a prerequisite for effective synchronization of operation and maintenance within the manufacturing industry as well as in other industries.

The relation between production plans and activities and actual production can to some extent be described deterministically. The relation between maintenance plans and activities and the production system availability on the other and requires probabilistic representation. The term *stochastic digital twin* is introduced whenever probabilistic models are required to represent the physical assets, processes and/or systems.

A wide range of mathematical models exist for predictive maintenance. In this presentation only two models will be presented to illustrate the relation between maintenance and production. Assume that we at a given point in time t_0 observe the state of a component. Let $\mathbf{y}$ be the vector describing the state at time t_0. Let T denote the point of time when the component fails. Let t denote running time from t_0. Given $\mathbf{y}$ assume that it is possible to describe the probability distribution of T. For simplicity we will as a starting point assume that T is Weibull distributed with shape parameter α and scale parameter λ. We assume that $\mathrm{E}(T|\mathbf{y})$ is relatively small, in the range days, weeks or months since we are in the condition monitoring situation. For example assume that we measure vibration in the bearing of a motor. An experienced maintenance engineer suggest that the mean time to failure, i.e., $\mathrm{E}(T|\mathbf{y})$ = 2 months = 60 days. When challenging him further and ask about the chances of failing earlier, after some discussions, he assess the probability that the component will fail within one month to be 10%. We will later on see how these assessment may be used to estimate α and λ.

Since the component is likely to fail rather soon it is obvious that a preventive maintenance activity should be conducted. For example replacing the bearing with a new one. The production manager, is however, not very happy with shutting down the production for maintenance. After rethinking, the production manager opens for using the weekly production shutdown for maintenance purpose. The first shutdown will come in $\tau_1 = 3$ days, the next in $\tau_1 + \tau$ days where $\tau = 7$ days, and so on. The cost of the preventive maintenance action will decrease the longer we can wait because planning may be improved. We assume a very simple cost structure where

$$c_{\text{PM}}(t) = c_{\text{PM},0}e^{-t/\theta} \tag{17}$$

where θ is a characteristic time for the decrease in the sense that for $t = \theta$ the cost has dropped to $e^{-1} \approx 37\%$.

The challenge is to determine which opportunity to apply, i.e., $t = \tau_1, \tau_1 + \tau, \tau_1 + 2\tau, \ldots$. The cost function to minimize is:

$$C(t) = c_{\text{PM},0}e^{-t/\theta} + c_{\text{U}}F_T(t) \tag{18}$$

Assuming that $\theta = 30$ days, $c_{\text{PM},0} = 10\ 000$ NOKs, and $c_{\text{U}} = 35\ 000$ NOKs we are just to start the minimization. The only challenge is to find the parameters α and λ.

For the Weibull distribution we have that $\text{E}(T) = \Gamma(1+1/\alpha)/\lambda$ and $F_T(t) = 1 - e^{-(\lambda t)^\alpha}$. Let x be such that $F_T(x) = p_x$ where both x and p_x are known. Further since $\text{E}(T)$ also is known we may in principle determine α and λ. A closed formula solution is not possible to obtain, but the following iteration scheme may be applied to obtain α:

$$\alpha_{i+1} = \frac{\ln(-\ln(1-p_x))}{\ln(x\Gamma(1+1/\alpha_i)/\text{E}(T))} \tag{19}$$

and then we resolve for the second parameter, i.e., $\lambda = \Gamma(1+1/\alpha)/\text{E}(T)$.

Example - Synchronization only age information only

For the example with $x = 30$ days, $p_x = 10\%$ and $\text{E}(T) = 60$ days we obtain $\alpha \approx 2.78$ and $\lambda \approx 0.0148$. Table 2 shows the result when applying Equation (18) for the calculations.

Table 2: PM-, unplanned failure (U)-, and total cost for the example

t	PM	U	Total
3	9 048	6	9 054
10	7 165	173	7 339
17	5 674	752	6 426
24	4 493	1 928	6 421
31	3 558	3 815	7 373

This means that the optimal time for performing the PM task will be the third or forth opportunity. □

Predictive maintenance and cyber physical systems

Predictive maintenance is about utilizing information regarding the condition of a component and the future expected loads in order to judge the correct time for intervention. In the previous section a simple model was derived but the current condition of the component and the future expected loads were not explicitly used. Some formalism is required for such a utilization. This will be crucial for cyber physical systems (CPS) where a computerized mathematical model of the system is established where real time information regarding state, production profile and plans etc are connected via IoT.

A reasonable simple extension of the model used in the previous section will be derived. The starting point is the failure rate function, $z(t)$. We stick to the Weibull distribution where the failure rate function is given by $z(t) = \alpha\lambda^{\alpha}t^{\alpha-1}$. We observe that $z(t)$ does not contain neither the current state nor the future loads. The so-called Cox-proportional hazard model is often used to incorporate the current state in the failure rate function. Let $\mathbf{y}$ be the vector of current relevant state information for the component, for example temperature, vibration level and so on. Next let $\overline{\mathbf{x}(t)}$ be the vector of average loads in the time period $[0,t)$. The failure rate function may be written on the form:

$$z(t|\mathbf{y},\overline{\mathbf{x}(t)}) = z_0(t)e^{\boldsymbol{\beta}_1\mathbf{y}}e^{\boldsymbol{\beta}_2\overline{\mathbf{x}(t)}} \tag{20}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are regression coefficient vectors established by for example statistical analysis of data. $z_0(t)$ is a baseline failure rate function, typically on the form $z_0(t) = \alpha\lambda^{\alpha}t^{\alpha-1}$

Now assume that the parameters α, λ, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are all known. Further assume that the current component state, $\mathbf{y}$, is known and that we have an estimate of future load $\overline{\mathbf{x}(t)}$. The cost equation to minimize is:

$$C(t) = c_{\mathrm{PM},0}e^{-t/\theta} + c_{\mathrm{U}}F_T(t|\mathbf{y},\overline{\mathbf{x}(t)}) \tag{21}$$

where the cumulative distribution function is given by:

$$F_T(t|\mathbf{y},\overline{\mathbf{x}(t)}) = 1 - \exp\left(-\int_0^t z(u|\mathbf{y},\overline{\mathbf{x}(u)})\,du\right) \tag{22}$$

A main objective for cyber physical systems is to set up a regime for data collection and analysis. It is beyond the scope of this presentation to describe relevant statistical methods. Typically a partial likelihood approach is recommended where the impact of the regression coefficient is estimated, and then a separate approach is used for estimation of the failure rate function.

If no data is available we might use expert judgements for elicitation of the relevant model parameters. The so-called PF-model is used as a basis for the elicitation. In the PF-model we assume that up to sompe point of time P there is no indication of a failure. But then starts failure progression until the failure progression exceeds the failure limit at point of time F. The point of time P is often referred to as a *potential failure* whereas the point of time F is a real failure. The time interval between the points P and F is denoted the PF-interval. The PF-interval is treated as a stochastic variable.

In the assessment $F_T(t|\mathbf{y},\overline{\mathbf{x}(t)})$ corresponds to the cumulative distribution function of the PF-interval. The procedure for the elicitation is as follows:

1. Assess the expected length of the PF-interval under the assumption of insignificant future load $\overline{\mathbf{x}(t)}$. Denote this value by ξ.

2. Asses the consistency of the PF-interval by the shape parameter α in the Weibull distribution. As a rule of thumb use

   - $\alpha = 2$ corresponds to a variety of failure mechanisms and causes leading to a failure.
   - $\alpha = 3$ corresponds to a few failure mechanisms and causes leading to a failure.
   - $\alpha = 4$ corresponds to a rather specific failure mechanism / failure cause leading to a failure.

3. Calculate the scale parameter by $\lambda = \Gamma(1/\alpha + 1)/\xi$.

4. For each y_i in $\mathbf{y}$ let $y_i = 0$ correspond to the condition at the point of time P in the PF-model. This corresponds to no significant degradation for the actual regression variable.

5. For each y_i in $\mathbf{y}$ let $y_{i,\mathrm{C}}$ be a critical value for that particular regression variable. Under the assumption that all other regression variables $y_j = 0, j \neq i$ assess the reduction in ξ by some factor, say k_i. Note that there is no specific "rule" to determine $y_{i,\mathrm{C}}$, and the higher value chosen, the lower value will be assessed for ξ.

6. Calculate the corresponding regression parameters by $\beta_i = -(\ln k_i)/y_{i,\mathrm{C}}$, i.e., for the elements in $\boldsymbol{\beta}_1$.

7. Repeat the procedure for each $\overline{x_i(t)}$ in $\overline{\mathbf{x}(t)}$ to find the elements of $\boldsymbol{\beta}_2$.

Using vibration data and expected average loads

The example above is now slightly modified to take into account explanatory variables. Let y be the vibration level measured by the so-called "RMS" value (Root Mean Square) which is an ISO convention. Technically the RMS value is calculated by multiplying the peak amplitude by 0.707. For machines of medium size the vibration level is mapped into zones where zone A is the normal level which we here assume corresponds to $y = 0$, zone B which still is considered acceptable ranges from $y = 1.8$ to $y = 4.5$, zone C which is critical ranges from $y = 4.5$ to $y = 11.2$ and zone D corresponds to $y > 11.2$. A machine in zone D is considered to have serious damages within very short time and is therefore often protected by a protection system causing the machine to shut down (TRIP). Further let x measure the portion of time the machine is run on more than 90% of maximum capacity.

In the elicitation process the maintenance engineer assess the mean residual time to failure, i.e., the time until the protection system will trip the system (PF-interval) to be $\xi = 120$ days when an anomaly situation occur, i.e., drifting into zone B. Since only vibration and excessive load is considered as influencing factors of the PF-interval the shape parameter is assessed by $\alpha = 4$. This gives $\lambda = \Gamma(1/\alpha + 1)/\xi = \Gamma(1.25)/120 \approx 0.00775$.

The critical value for the vibration is set to $y_C = 4.5$ and the corresponding reduction factor for the remaining time to trip is assessed to $k_Y = 0.05$ (only 6 days to failure in average). This gives $\beta_Y = -(\ln k_Y)/y_C = -(\ln 0.05)/4.5 \approx 0.666$.

A machine running with 90% of maximum capacity or more in $x_C = 0.25 =$ 25% of the time is assessed to have a reduction factor of $k_X = 0.1$ (12 days to failure in average). This gives $\beta_X = -(\ln k_X)/x_C = -(\ln 0.1)/.25 \approx 9.2$.

The relevant parameters to calculate the cumulative distribution function for the PF-interval in Equation (22) have now been established. Now assume that we have observed $y = 3$ and we assess future loads to be $x = 0.1$. Inserting in Equation (22) and using the cost function in Equation (21) Table 3 indicates that we should use the opportunity that comes after 31 days.

Table 3: Results

t	PM	U	Total
3	13573	0	13573
10	10748	21	10769
17	8511	176	8687
24	6740	693	7433
31	5337	1894	7231
38	4227	4132	8358

Note that the cumulative distribution function calculated by Equation (22) is the unconditional distribution function given we were at point of time P in Figure **??**. In reality since $y = 3$ it is reasonable to believe that some days has elapsed since the potential failure was evident. A conditional distribution function is therefore more appropriate. This means that we also need to assess the time since the potential failure occurred. Let t_0 be the current time, and assume that the time since the potential failure (P) is s time units. Let t denote the running time from now on, i.e., t_0 corresponds to $t = 0$. Using the rule for conditional probabilities we obtain the following modified cost function:

$$C(t) = c_{\text{PM},0}e^{-t/\theta} + c_{\text{U}}\left[1 - \frac{1 - F_T(t+s|\mathbf{y}, \overline{\mathbf{x}(t+s)})}{1 - F_T(s|\mathbf{y}, \overline{\mathbf{x}(s)})}\right] \tag{23}$$

In the example calculation this conditional approach is not used. □

Example - Towards a real time model - The digital twin

The example is now used as motivation for developing a simple stochastic digital twin. A digital twin may be viewed as a digital simulation model with built in analytics, decision support, and self learning features. Learning features will not be discussed in this example, and only glimpse of analytics is provided.

The digital twin is represented by two models, one maintenance model and one production model, where these models interact via the Internet of Things. In the following the maintenance model is denoted the maintenance twin and the production model is denoted the production twin. The physical counterpart of the maintenance twin is the actual component state, the physical load on the machine, the actual maintenance carried out the actual time the machine can not produce due to preventive and/or corrective maintenance and so on. The physical counterpart of the production twin is what is actually being produced, when the production takes place, the economic value of the production, the cost of production, the various machines being used, the use of personnel and resources and so on.

Let $\mathscr{T}$ be the operational windows for execution of a preventive maintenance task of the packing machine, i.e., the point of times $\tau_1, \tau_1+\tau, \tau_1+2\tau, \ldots$. The decision support to be provided by the maintenance twin upon a potential failure situation is now:

$$\min_{t\in\mathscr{T}} C(t) = c_{\text{PM},0}e^{-t/\theta} + c_{\text{U}}F_T(t|\mathbf{y}, \overline{\mathbf{x}(t)} \tag{24}$$

The maintenance twin represented by Equation 24 has to be implemented on a digital platform, for example MS Excel. The maintenance twin needs to be fed with data from the production twin. Here the production twin is very simple, only a set of predefined scenarios combining different values of $c_{\text{PM},0}, c_{\text{U}}, \mathbf{y}$, and $\overline{\mathbf{x}(t)}$. Table 4 shows the data used in this simple MS Excel representation of the two twins interacting. In a real life implementation the data in Table 4 needs to be generated by the ERP system, the SCADA system and so on.

In order to communicate with the maintenance twin, the production twin needs to post data to the internet

Table 4: Data used in the production twin

$c_{\mathrm{PM},0}$	c_{U}	y	x	CPS message/Comment
15000	35000	3	0.1	Base line (from example)
15000	35000	3	0.3	High future loads
5000	35000	3	0.1	Cheap PM due to low production
15000	35000	2	0.15	Lower degradation
15000	35000	4	0.15	Very high degradation
15000	100000	3	0.15	Very high failure cost

```
Function postInternetData(data)
' Make the data available on internet
End Function
```

Similarly, the maintenance twin needs to get access to this data:

```
Function getInternetData(specification)
' Get data from internet with some specification of what to get
End Function
```

The maintenance twin is continuously reading data from the internet to come up with maintenance decision support. This information should then be posted on the internet and incorporated in the production twin for production and maintenance planning. This has not been implemented in this very first example. □