On the effect of downtime costs and budget constraint on preventive and replacement policies

R. Pascual\textsuperscript{a,}\textsuperscript{*}, V. Meruane\textsuperscript{a}, P.A. Rey\textsuperscript{b}

\textsuperscript{a}Department of Mechanical Engineering, Universidad de Chile, Casilla 2777, Santiago, Chile
\textsuperscript{b}Department of Industrial Engineering, Universidad de Chile, Casilla 2777, Santiago, Chile

Received 10 November 2005; received in revised form 29 September 2006; accepted 6 December 2006
Available online 12 January 2007

Abstract

This work proposes a general approach to study and improve the effectiveness of the system with respect to its expected life-cycle cost rate. The model we propose considers a production system which is protected against demand fluctuations and failure occurrences with elements like stock piles, line and equipment redundancy, and the use of alternative production methods. These design policies allow to keep or minimize the effect on the nominal throughput, while corrective measures are taken. The system is also subject to an aging process which depends on the frequency and quality of preventive actions. Making decisions is difficult because of discontinuities in intervention and downtime costs and the limited budget. We present a non-linear mixed integer formulation that minimizes the expected overall cost rate with respect to repair, overhaul and replacement times and the overhaul improvement factor proposed in the literature. The model is deterministic and considers minimal repairs and imperfect overhauls. We illustrate its application with a case based on a known benchmark example.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Decision-making; Downtime cost; Maintenance policies; Repair rate; Replacement; Imperfect maintenance; Decision support systems; Consequential cost; Budget constraint

1. Introduction

When production equipment fails or is overhauled, the costs that the company pays may be classified in two categories. In a first category, we consider the intervention costs, which include labor and materials. In the second category, we include downtime costs, which consist of the cost of lost production as well as other consequential costs, such as reconfiguring alternative production lines, using less efficient methods, reduced product quality, lost raw material, and so on [1]. Quantifying intervention costs is quite straightforward since standard accounting procedures register them. On the other hand, downtime costs may be hard to estimate because they depend on several external factors like production rates, stock prices, and system design parameters (redundant equipment, stock piles, and alternative production methods). The sum of both intervention and downtime costs define the global cost, which is a very interesting indicator to measure the success of maintenance management strategies [2].

Estimating downtime costs can be of much benefit to maintenance decision-making for several reasons: (1) it allows to measure the impact of equipment on system efficiency; (2) it may be used to assess the effectiveness of maintenance policies (key performance indicator); (3) it allows the use of a series of mathematical models applied in decision-making contexts, such as replacement policies, preventive and condition-centered maintenance strategies, and spares stock levels.

This article reviews the effects of downtime costs and a budget constraint on maintenance decision-making. For illustrative purposes, we will assume that cost curves have been previously estimated. We focus on the effect of discontinuities on the decision-making process related to setting overhaul times and quality, service and replacement times.
In the next section, we present a literature review. Then, we state the problem and introduce the model in Section 3. We discuss the failure rate model used in Section 4. In Section 5, we analyze an example derived from Example 1 from Zhang and Jardine [3]. Finally, we comment on the results and provide possible lines of further research.

2. Literature review

A significant amount of research in the area of reliability engineering considers the study of maintenance policies to prevent system failures and improve system availability. Maintenance modeling has been recognized as a growing subject and also as a late developer [4]. This situation is partially explained because maintenance take actions on plant equipment and not on product. As a consequence, it is perceived as a marginal activity of the company. Scarf [4], Dekker and Scarf [5] and Wang [6] present very good reviews in the area of maintenance models, assessment of their impact and applications.

Maintenance decision-making concentrates in general on costs, thus, it is necessary to analyze them. Komonen [7,8] proposes a classification of costs related to maintenance considering two groups: (1) direct (intervention) costs; due to maintenance operations (administrative costs, labor, material, subcontracting), (2) lost production costs due to equipment failures and lost quality production due to equipment malfunctioning. This classification emphasizes two conflicting maintenance objectives: maximizing equipment availability while reducing intervention costs.

Traditional cost models assume that lost production dominates the downtime cost and neglect discontinuities due to stock piles exhaustion, lost raw material, etc. These models assume constant production rates and unit prices, then downtime costs rise linearly with maintenance service time (see [9]). Vorster and De la Garza [1] segregate different sources for the consequential costs and propose a non-linear time-dependency of downtime and intervention costs as it is illustrated in Figs. 1 and 2, respectively. They classify the consequential costs when a machine breaks down in four sub-component costs associated with lack of availability and downtime. Associated resource impact costs concern the effects of the failure on other components of the team. Lack of readiness costs are defined as the penalty costs that could be levied due to the expectation that resources representing capital investments in productive assets be kept in an operational condition most of the time. Service level impact costs measure the decreased productivity of a fleet of equipment when a portion of that fleet has failed. Alternative method impact costs occur when a different method of production must be used due to the failure of a given component of the original production team. Mitchell [10] uses regression analysis to represent equipment costs in terms of age. After identifying the model parameters, a cumulative cost model is used to identify optimum economic decisions for repair policies, and overhaul and replacement intervals.

Analytical results to quantify downtime costs are generally difficult to obtain. For example, Roman and Daneshmend [11] use simulation to study the effect of contractors on service level in open-pit mines. Quintana and Ortiz [12] consider simulation for resource assignment in a maintenance shop. Cor [13] uses a similar strategy to quantify the time and economic impacts of operational changes. These changes may include alterations to the sequence of work, to the design, or non-stationary conditions. As a consequence, they may influence the quantity and type of resources required to perform the works, and logically, on equipment idle time and usage. A supplementary strategy considers the use of past information: Edwards and Yisa [14] and Edwards et al.
[15] use regression analysis to predict the expected downtime cost rate for tracked hydraulic excavators in open cast mining industry. Their metamodel shows that in that case, machine weight is an excellent predictor of downtime cost.

Vorster and Sears [16] propose what they call failure cost profiles which measure the expected cost per unit time in terms of the duration of the downtime interval. For a fixed repair time, they introduce a cost-related criterion that also takes into account the relative productivity of multifunctional equipment. By doing so, they are able to decide replacement and task assignments for a fleet of similar equipment.

Operator-induced consequential costs are studied by Edwards et al. [17]. The work includes skill level of operators, fatigue, morale, and motivation. Operator’s skill is one of the most important factors as it can have a great impact on performance of the equipment. Considering operational poor practices, Pathmanathan [18] reports the increased frequency and cost of equipment downtime induced by the negligence of the operator and lack of proper training and knowhow on the part of the equipment supervisor. Arditti et al. [19] also consider operation uncertainty (i.e., different environmental conditions), and design complexity as causes of greater risk for equipment downtime. Nepal and Park [20] explore the impact of downtime on construction projects duration and related costs. Their analysis highlights how various factors and processes interact with each other to cause downtime, and mitigate or exacerbate its impact on project performance.

As for preventive maintenance policies, Kobayashi et al. [21,22] consider decisions on multi-component systems. The issue is to determine the optimal maintenance preventive times to minimize total downtime due to both corrective and preventive interventions for a large set of components. A decision support system is used to carry out an automated analysis of the maintenance history data. Maintenance data are input to the system and the most suitable mathematical model from a model-base is identified. Another approach considers the use of expert systems for maintenance project planning. Moynihan et al. [23] apply the concept to an automotive components manufacturer.

To the cost categories previously listed we may add other two, which we will consider constant in our model: spare holding costs and costs associated to investment on assets that are based on maintenance criteria like system reliability, availability and maintainability (see [2]).

Pham and Wang [24] review the literature on imperfect maintenance. They classified maintenance works according to the degree to which the operating condition of an item is restored by maintenance as: (a) perfect maintenance: the system is restored to an operating condition as good as new and has the same lifetime distribution and failure rate function as a brand new one; (b) minimal maintenance: the system is restored but the failure rate is the same rate it had just prior to the failure; (c) imperfect maintenance: the system is not as good as new, but younger than it was at failure time; (d) worse maintenance: the system failure rate or actual age increases but the system does not break down. As a result, the operating condition becomes worse than that just prior to the intervention; (e) worst maintenance: a maintenance action which undeliberately makes the system fail. The failure/repair process for repairable items is also well explained by Calabria et al. [25]. Martorell et al. [26] considered an equipment age-dependent reliability model for nuclear power plant operation. In particular, they presented the results of a sensitivity study, and conclude that maintenance effectiveness has a major impact on the selection of an appropriate maintenance strategy.

3. Statement of the problem

Let us consider a situation where an equipment may suffer failures in time according to a distribution with probability density function \( f(t) \). Corrective actions are applied when the equipment fails. It also may be subject to periodic overhauls. After some time \( T_{c1} \), the equipment is replaced by a new one. Repairs and overhauls take \( T_{r} \) and \( T_{o} \) time units to be performed, respectively, and the intervention costs are \( C_{r} \) and \( C_{o} \) money units (\( \text{mu} \)), respectively. Maintenance decision-makers decide on these values in search of maximum profitability for the company. Downtime costs also depend on the maintenance service times. Some discontinuities appear (see Fig. 1) on the curve due to the exhaustion of stock piles, use of alternative, less efficient methods, and so on. In general this function is non-decreasing in \( T_{r} \) and \( T_{o} \). Intervention costs may also be estimated from \( T_{c1} \). We will consider, for the same set of maintenance actions, that the intervention cost is a non-increasing function of \( T_{c1} \) (Fig. 2) and may also be discontinuous due to scale economies, use of extra resources and more expensive repair methods. There could be segments where maintenance actions are not possible (i.e., it takes at least \( T_{c_{min}} \) time units to repair). The linking of the different variables in the decision-making process is shown in the concept map of Fig. 3 (see also [27]).

We make the following assumptions:

1. Corrective actions are \( \text{minimal} \).
2. Preventive actions are applied at equal time intervals \( T_{c1} \).
3. Opportunistic maintenance practices may be disregarded due to the lack of scale economies with grouped maintenance actions [28,29].
4. Life-cycle duration is equivalent to an integer \( n \) number of periods between overhauls. That is, \( T_{c1} = n T_{c_{min}} \). Consequently, there are \( n - 1 \) overhauls during a life-cycle.
5. Intervention costs depend on the time interval required to perform the activities. They correspond to piecewise linear functions. Each segment is determined by \( T_{c_{ij}} \) and costs \( C_{rij} \) and \( C_{rij} \) on the left and right side of segment \( j \).
(6) Overall replacement cost corresponds to $C_{R,g}$:
\[ C_{R,g} = C_{R,i} + C_{R,d}. \]

(7) The objective is to minimize the expected overall cost rate in order to determine the overhaul frequency and maintenance service time. The choice is justified as it is the only one that explicitly measures the economic effects of downtime costs.

Another objective considered in the literature for similar problems includes, among others:

- To maximize availability: this criterion is appropriate when downtime costs are much larger than intervention costs. It applies when considering critical equipment [2,9].
- To minimize intervention costs: once availability is modeled, it is direct to estimate intervention costs and to fix a budget. This criterion does not take explicitly into account downtime costs. In order to this, it is necessary to add a minimum availability constraint explicitly. This would be achieved imposing a minimum availability constraint [3,11,28].

The difficulty of using the overall cost rate as objective comes from the identification of downtime costs, which may be difficult to estimate. The amount of information needed to use this objective is much larger than using others. Obtaining the required data from the information system could be difficult or costly.

3.1. Statement of the model

In what follows, subscripts \{o, r, R\} refer to overhaul, repair, and replacement, respectively (they also appear as superscripts). Intervention type related subscripts are \{i, d\} (intervention, downtime). Superscripts l and r refer to left and right of a segment, respectively. Subscript j refers to segment j. In order to simplify model formulation, we will consider that segments are defined for both intervention and downtime costs; and that corrective and preventive intervention costs have ratio $\kappa$ for any given duration.

First, we introduce decision variables, then some auxiliary variables to simplify the description of the model and finally, we state the constraints and objective function. The budget constraint is discussed at a separated paragraph.

3.1.1. Decision variables

- $T_l$, life-cycle duration;
- $T_s$, time interval between two preventive interventions;
- $T_o$, time interval required for a preventive action;
- $T_r$, time interval required for a repair.

3.1.2. Auxiliary variables

- $C_{o,i}$, preventive intervention cost;
- $C_{o,d}$, downtime cost for overhauls;
- $C_{c,i}$, corrective intervention cost;
- $C_{c,d}$, downtime cost for repairs;
- $C_{o,g}$, global cost of an overhaul (intervention plus downtime costs);
• $C_{>,g}$, global cost of a repair (intervention plus downtime costs);
• $T_{o,j}$ equals $T_o$ if it lies in the $j$th interval, otherwise it is 0 ($j = 1, \ldots, J$);
• $T_{r,j}$ equals $T_r$ if it lies in the $j$th interval, otherwise it is 0 ($j = 1, \ldots, J$);
• $\delta^o_j$, binary variable which is 1 if $T_o$ lies in the $j$th interval ($j = 1, \ldots, J$);
• $\delta^r_j$, binary variable which is 1 if $T_r$ lies in the $j$th interval ($j = 1, \ldots, J$);
• $n$, number of failures during the life-cycle;
• $n$, number of overhauls during the life-cycle minus one.

Note that $T_{r,j} = T_o \delta^o_j$.

3.1.3. Constraints

• In order to make $T_{o,j}$ to be active only in the corresponding interval:
  \[ \delta^o_j T^1_{o,j} \leq T_{o,j} \quad \text{for } j = 1, \ldots, J, \]

\[ T_{o,j} \leq \delta^o_j T^1_{o,j+1} \quad \text{for } j = 1, \ldots, J \] (1)

• and for repairs:
  \[ \delta^r_j T^1_{r,j} \leq T_{r,j} \quad \text{for } j = 1, \ldots, J, \]

\[ T_{r,j} \leq \delta^r_j T^1_{r,j+1} \quad \text{for } j = 1, \ldots, J. \] (2)

• Computing downtime costs:
  \[ C_{o,d} = \sum_j \left[ C^d_{d,j} + \frac{C^r_{d,j} - C^d_{o,j}}{T^1_{o,j+1} - T^1_{o,j}} (T_{o,j} - T^1_{o,j}) \right] \delta^o_j, \]

\[ C_{r,d} = \sum_j \left[ C^d_{d,j} + \frac{C^r_{d,j} - C^d_{o,j}}{T^1_{r,j+1} - T^1_{r,j}} (T_{r,j} - T^1_{r,j}) \right] \delta^r_j. \]

• Computing intervention costs:
  \[ C_{o,i} = \sum_j \left[ C^i_{o,j} + \frac{C^r_{o,j} - C^i_{o,j}}{T^1_{o,j+1} - T^1_{o,j}} (T_{o,j} - T^1_{o,j}) \right] \delta^o_j, \]

\[ C_{r,i} = \kappa \sum_j \left[ C^i_{d,j} + \frac{C^r_{d,j} - C^i_{r,j}}{T^1_{r,j+1} - T^1_{r,j}} (T_{r,j} - T^1_{r,j}) \right] \delta^r_j. \]

• Only one active interval:
  \[ \sum_j \delta^o_j = 1, \] (7)

\[ \sum_j \delta^r_j = 1. \] (8)

• $\delta^o_j$ and $\delta^r_j$ are binary variables for all intervals $j$; $n$ and $n_r$ are positive integer variables. Remaining variables are continuous and non-negative.

3.1.4. Objective function

The expected global cost per unit time for this model may be expressed as follows:

\[ c_g(T_s, T_o, T_r, n) = n_s(nT_s)(C_{c,d} + C_{c,i}) + (n - 1)(C_{o,d} + C_{o,i}) + C_{R,g}. \] (9)

3.1.5. Budget constraint

Fig. 4 represents the long term trend of intervention costs. We observe how the increasing failure rate produces more repairs as time passes. The situation may absorb the scarce resources allocated to maintenance. To handle that, let us consider that a constant budget $B$ is allocated at the beginning of a given strategic time unit (i.e., each year).

The worst situation appears at the last part of the life-cycle and specifically during the last year. As the budget is constant, the budget constraint limits the allowable number of failures during the last year (note that the life-cycle corresponds to the multiplication of two decision variables, $n$ and $T_s$):

\[ C_{ir} \int_{nT_s - k}^{nT_s} \lambda \, dt + C_{io} n_{o,i} \leq B, \] (10)

where $n_{o,i}$ corresponds to the number of overhauls during the last year. As the budget constraint holds for the last year, previous budgets are not fully used unless they are applied for preventive actions which reduce the failure rate afterwards, and specifically during the last year.

In the next section we will describe the model used to compute the failure rate function.
4. Failure rate model

Zhang and Jardine [3] propose a failure rate model to describe the situation where the equipment receives imperfect preventive maintenance actions at discrete moments in time, with a fixed frequency. Corrective actions are minimal in their model.

In the model considered here, after an overhaul, the equipment is between as good as before and as good as after previous overhaul. Let $\lambda_k(t)$ be the failure rate after the $k$th overhaul. The failure rate will be expressed as

$$\lambda_0(t) = \tilde{\lambda}(t),$$

$$\lambda_k(t) = p\lambda_{k-1}(t - T_s) + (1 - p)\lambda_{k-1}(t), \quad k = 1, \ldots, n - 1,$$

where $p \in (0, 1)$ corresponds to an improvement factor that models the degree of imperfection of the overhaul; and $\tilde{\lambda}(t)$ is a baseline failure rate function that models the situation before the first overhaul (and determines the behavior of $\lambda_k(t), \quad k = 1, \ldots, n - 1$). The failure rate function is discontinuous. Fig. 5 illustrates its behavior.

The curve with no discontinuities represents an equivalent for $\tilde{\lambda}(t)$ in terms of the expected number of failures and its parameters are estimated in Pascual and Ortega [30].

5. Numerical example

We will now consider the example of Zhang and Jardine. We modify the conditions of the case to consider the effect of an annual budget constraint $B$ and the times to repair and overhaul the equipment ($T_r$ and $T_o$, respectively). In order to reduce the search space and take into account that strategic decisions are made on an annual basis we consider

$$T_s = \frac{365}{v} \text{ days},$$

$$T_i = 365\kappa \text{ days},$$

where $v$ and $\kappa$ are positive integers. Thus, decision variables in the analysis of this section are $T_r, T_o, v$ and $\kappa$.

Intervention and downtime costs decrease with the time interval to perform them as it is observed in Fig. 6. These costs vary linearly on each segment according to data provided in Table 1.

It can be noticed that during the first time units no downtime costs exists. This can be explained by the existence of contingency measures like stock piles and alternative equipment. Combining both intervention and downtime cost we obtain the profiles of global costs for both corrective and preventive interventions (Fig. 7). We notice that minimal global costs are $2 \text{mu}$ for the case of corrective actions (at $T_r = 2$ days) and to overhaul the equipment in 14 days. If budget is large enough, the maintenance department would repair and overhaul the equipment according to these times to maximize company’s profitability.

If we set a life-cycle of 6 years and 2 overhauls/year ($\kappa = 6$ and $v = 2$) and no budget constraint applies, we

![Fig. 5. Model from Zhang and Jardine vs. model proposed (from the example, $p = 0.7$).](image)

![Fig. 6. Expected costs (mu/action).](image)

<table>
<thead>
<tr>
<th>$T_r$</th>
<th>$C^l_t$</th>
<th>$C^l_r$</th>
<th>$C^l_{io}$</th>
<th>$C^l_{ir}$</th>
<th>$C^l_{io}$</th>
<th>$C^l_{ir}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>12</td>
<td>35</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>2</td>
<td>12</td>
<td>6</td>
<td>1.8</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>100</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>90</td>
<td>100</td>
<td>1000</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Cost parameters
obtain the cost surface shown in Fig. 8. We corroborate that the best decision for the company is to repair and overhaul at $T_r = 2$ days and $T_o = 14$ days at an expected global cost of 50.44 mu/year. Fig. 8 also superimposes the budget constraint (labeled hyperbola-like curves) over the isocost lines of the expected global cost per unit time.

Fig. 9 summarizes the budget sensitivity analysis. Life ranges between 3 and 6 years. The budget constraint is active if the budget available is 25 mu/year or lower. Below 9 mu/year $T_o$ has to be incremented. $T_r$ is more sensitive and shows discontinuities for different life-cycle durations ($T_l = 3, 4, 5, 6$ years).

5.1. Discussion

Having Figs. 8 and 9, facilitate decisions related to replacement cycle, adequate preventive and corrective programs, size of maintenance work force and outsourcing of services. If, for example, a company is replacing the equipment every 6 years, and overhauling every 6 months (Fig. 8) with a duration of overhaul of 50 days, global maintenance costs may be cut by a factor of 3 simply by reducing the duration of overhauls to 15 days even if it costs more in terms of intervention costs as outsourcing or enlargement of the work force would be required. A strategic view of the situation is clearly gained. The short-sighted decision-making according to the annual budget constraint is avoided.

Determining optimum intervention times would allow also to assess the marginal benefits/costs of alternative production schemes (safety stock, redundant equipment, etc.) and maintenance outsourcing contracts.

In an strategic decision setting, results and analysis obtained could be used to evaluate and improve the design.
of the production line, modifying contingency measures if appropriate.

6. Conclusions and further work

Real industrial systems almost always have design alternatives to handle load fluctuations and failure occurrence. This situation implies discontinuities in the cost functions that measure the impact of maintenance policies. This work develops an approach to address the problem through the use of a non-linear mixed integer programming formulation. This type of models is computationally difficult to solve since it combines both the non-convex nature of the global cost with the combinatorial nature of integer variables. The development of algorithms for mixed-integer non-linear models is an active research area. We are working on adapting recent developments (see for instance [31,32]) to solve the model proposed.

A numerical example has shown the feasibility of improving maintenance decision support systems to take into account the effect of palliative measures in real industrial situations. We have studied a situation where \( T_r \) and \( T_o \) are to be fixed in order to maximize profits. Further developments could allow it change according to observed failure rates and task-oriented and time-dependent budgeting. Multi-objective optimization of the above decision variables is also an interesting issue for further research as it allows considering criteria such as failure risk, availability, etc.

Acknowledgments

The authors wish to acknowledge the partial financial support of this study by the Millennium Science Nucleus Complex Engineering Systems F-06-04. We would also like to thank the reviewers for their constructive observations.

References