Assessing Uncertainty from Data Collection to Maintenance Optimization

Guillaume Fleurquin, Christophe Letot and Pierre Dehombreux
Risks Research Center
Faculté Polytechnique de Mons
Rue du Joncquois, 53 – B7000 Mons (Belgium)
guillaume.fleurquin@fpms.ac.be;
christophe.letot@fpms.ac.be;
pierre.dehombreux@fpms.ac.be

Fouad Riane
Center of Research and Studies in Industrial Management
Catholic University of Mons
Chaussée de Binche 151 – B7000 Mons (Belgium)
fouad.riane@fpms.ac.be

Abstract— In this paper, we will propose a framework to perform the optimization of periodical maintenance tasks for a production line, with a specific viewpoint on uncertainty issues from the modelling step to the analysis of numerical results. From a structured log file of operational data, we build a reliability-based model (block diagram) that is used to optimize the parameters of the maintenance policies through Monte Carlo simulations. The model is determined by using every data source available (Computerized Maintenance Management System hierarchy and failure mode classification especially).

Keywords- intelligent maintenance, Monte Carlo simulation, optimization, uncertainty.

I. INTRODUCTION

Maintenance is one of the fields in which societies can increase their performance: production costs, availability, quality, security… But it is also one of the most difficult to master since it relies on knowledge (intrinsic data of equipments) and experience (operational databases). Moreover it is strongly subject to uncertainty for many kinds of equipments. As a consequence, a reliability study must use all the available data sources, as far as they convey an informational content. This is a key point for intelligent maintenance for which all kinds of data related to the degradation of the equipment (CMMS, quality control data, process parameters) are gathered. Otherwise results will be far too imprecise to allow maintenance managers to exploit them.

Many contributions can be found upon reliability of repairable or non repairable systems [7], [3], [8]. Different kinds of maintenance may be considered to optimize performances: corrective, preventive or conditional. In this paper, we will focus on time-based preventive maintenance. Since the Age Replacement Policy [1], many strategies have been elaborated. Two other well-known strategies are the Block Replacement Policy [6] and its variant the Modified BRP. In [9] and [4], we can find a large panel of these variations of maintenance policies.

In order to conduct various industrial studies, our department developed an adaptable framework and numerical tools. Indeed, industrial databases are often quite different depending on their application domains. The pre-processing of these data is so crucial. We focus our studies on maintenance activities that are time oriented. Last, we can find in [2] additional complements on uncertainty in reliability analysis.

II. METHODOLOGY

To make the analysis of complex systems, simulation based on a reliability block diagram allows absorbing most of the aspects of these systems and composing with their complexity. Such a diagram is based on the association in series or in parallel of entities which represent either physical components or failure classes. In the former case, the decomposition is made in agreement with the structure in the CMMS. In the latter case, components are grouped, in an operational perspective, into families of failure and repair classes (hydraulic, electrical...). Then the economic data have to be collected, which includes production and maintenance ones.

Operational data are used to extract reliability laws for each component of the model. A selection between all the parts of the system has to be made, because it cannot entirely be modelled, thanks to a Pareto analysis for example. Furthermore, we need a flexible data structure that is adapted to every industrial system and contains all we need. Table 1 shows the file format we have established and some examples of entries. First, time must be measured with precision. Then, for each event, the state of the component must be defined:

- Working. The component is running.
- Stand-by. The component is able to run but the system is stopped due to an external cause.
- Failure. The component has failed (and is not yet being repaired).
- Corrective maintenance. The maintenance team is working on the component.
TABLE I. FILE FORMAT FOR COLLECTING OPERATIONAL DATA

<table>
<thead>
<tr>
<th>Date / Time</th>
<th>State</th>
<th>Failure mode</th>
<th>Serial number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>06/24/2008-18:56</td>
<td>Corrective maintenance</td>
<td>Mechanical</td>
<td>02.225</td>
<td>Oven: small crack on surface</td>
</tr>
<tr>
<td>06/25/2008-06:30</td>
<td>Running</td>
<td>-</td>
<td>02.225</td>
<td>End of repair</td>
</tr>
<tr>
<td>06/28/2008-22:00</td>
<td>Stand-by</td>
<td>External</td>
<td>-</td>
<td>Power cut</td>
</tr>
<tr>
<td>06/28/2008-22:00</td>
<td>Running</td>
<td>-</td>
<td>-</td>
<td>End of power cut</td>
</tr>
<tr>
<td>06/30/2008-13:12</td>
<td>Failure</td>
<td>Hydraulic</td>
<td>05.105</td>
<td>Joint broken</td>
</tr>
<tr>
<td>06/30/2008-13:34</td>
<td>Corrective maintenance</td>
<td>Hydraulic</td>
<td>05.105</td>
<td>Joint broken - repair</td>
</tr>
<tr>
<td>06/30/2008-14:15</td>
<td>Running</td>
<td>-</td>
<td>05.105</td>
<td>End of repair</td>
</tr>
<tr>
<td>07/01/2008-22:00</td>
<td>Preventive maintenance</td>
<td>-</td>
<td>-</td>
<td>Major maintenance</td>
</tr>
</tbody>
</table>

- Preventive maintenance. The maintenance team makes the required activities (replacements...).

The “failure mode” indicates the failure class. The “serial number” is necessary to identify the component and the “description” is the precise expression of the event.

The simulation model is based on Monte Carlo simulations. The results are the performances – depending on the objectives – under a specific maintenance policy. The parameters of this policy (typically, the preventive period of each component of the model) are then chosen to minimize (or maximize) the objective(s).

III. RELIABILITY ANALYSIS: STEP BY STEP

A. Structure of the system and modelling

When facing a new production line, one of the first questions is the degree of accuracy of the databases, that is to say until which level of decomposition there is data. A detailed mechanical structure of the line often exists, but operational data do not always reach the same degree of accuracy. The limit of the accuracy of the reliability study is the limit of the operational data (or the classification). Depending on the production line and on the machines, a single reliability distribution may well describe a component (Weibull distribution often fit the data). But if the latter is made of highly different units (for example, a high-frequency welder associated with a conveyor belt in the CMMS), the block should be considered as the sum of different failure modes. After such an analyse, we obtain the backbone of the block diagram.

Apart from the logical structure of the production line, a “failure mode classification” may be deduced from a log file like the one in table 1. This parallel classification completes the physical decomposition as shown by Fig. 1. When possible, a more accurate tree model is built: machines are physically divided into units that can be modelled as the association of different failures modes.

B. Identification: some considerations

For each component (or each failure mode), a reliability distribution has to be identified, its life must be rebuilt and each event that occur to it must be classified. Using the log file created, reliability laws are identified among the following ones: exponential, normal, lognormal, Weibull, extreme values and mixed Weibull [5]. Here are some of the characteristics of the latter.

The mixed Weibull distribution is defined as a sum of classical Weibull distributions, weighted by a factor that represents the proportion of the sub-populations to which applies each classical Weibull law. Its reliability function is defined as

\[
R(t) = \sum_{i=1}^{S} \frac{N_i}{N} \exp \left( - \left( \frac{t}{\eta_i} \right)^{\beta_i} \right)
\]

where S is the number of sub-populations, N is the total number of data and Ni and (\( \beta_i, \eta_i \)) respectively the number of data and the parameters of the Weibull distribution of the ith sub-population.

The identification of real data is often uneasy because of the complexity of some components. A component regroups indeed often more than one failure mode, and the identification of a single law may be inadequate. In these cases, the use of the mixed Weibull distribution should be a solution.

C. Optimization of maintenance activities

A multi-component system is often too complex to easily allow the optimization of the multiple maintenance activities of its components. We have chosen the way of simulation to estimate performances of some strategies, which helps choosing parameters to fulfil the goals of the maintenance managers. Besides, simulation allows considering uncertainty on models, but its inherent variability tends to reproduce the variability of reality. That is to say that there are two points of view: the uncertainty on models express the limits of the statistical analysis and give limit values of performances. On another side, variability of simulation expresses the risk inherent to unavoidable randomness of many failures.
IV. CASE STUDY: PHARMACEUTICAL PRODUCTION LINE

In the pharmaceutical domain, quality is synonymous to health security and is so a crucial indicator. As a consequence, each doubt upon the quality of a product leads to a stop of the production line and an inspection (or repair if a failure is effectively detected). The failed item(s) is (are) discarded (or recycled if possible). We will consider a production line of pouches. This line is supervised by an operator and the products are visually checked. This equipment is connected to the CMMS which save all events with a precision of one second. Maintenance teams have decided on a classification of all the events that we will describe below. The objective of this study is not economic: production must approach perfection, and maintenance is one of the bricks to achieve the quality goals.

The events that occur on the production line are not only due to failures. Calibrations or modifications of the running speed are also events to save. The first level of the classification makes notably the difference between intrinsic and external causes of stops. The second level localizes the component or the process that undergoes the event.

A. Reliability analysis: identification

We want to make a reliability analysis on the major kinds of failures. According to the Pareto principle we select the failures that produce the longer stop durations. For each of them, non parametric reliability laws show inflexion points. Fig. 3, which is our interface for the study of multi-Weibull distributions, shows an example of linearized non parametric Weibull distribution:

$$\log \left( \log \left( \frac{1}{1 - F(t)} \right) \right) = f(\log(t)) \quad (2)$$

Table 2 shows the results of the identification of the class “generic” for the classical Weibull and the multi-Weibull distributions. The three sub-populations for the multi-Weibull have been graphically determined, as explained earlier. The confidence intervals of each parameter are also given.

The difficult work of identification is well illustrated here: the identified parameters of the classic Weibull distribution indicate a tendency to youth failures ($\beta < 1$) but the multi-Weibull clearly shows that there is a first sub-population that comes from micro-stops and does not really underline real failures. The two other sub-populations indicate two sub-components which are subject to degradation.

When inflexion points clearly appear, like in Fig.2, a reflexion upon failure modes has to be introduced. Indeed, some defaults in the settings of the equipment may lead to multiple “failures”: some sensors could send successive alarms because of repetitive misplacement of raw material, for example. In that case, regrouping some points of a temporal point of view may lead to a more realistic model. Besides, different failure modes (hydraulic etc.) may be the origin of the different slopes of the curve.

<table>
<thead>
<tr>
<th>Class</th>
<th>Distribution</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sub-population 1</td>
<td>sub-population 2</td>
</tr>
<tr>
<td>generic</td>
<td>multi proportion</td>
<td>23.8%</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.49 &lt; 0.51 &lt; 0.54</td>
</tr>
<tr>
<td></td>
<td>$\eta$</td>
<td>32.9 &lt; 37.7 &lt; 43.1</td>
</tr>
<tr>
<td>classic</td>
<td>$\beta$</td>
<td>0.52 &lt; 0.54 &lt; 0.55</td>
</tr>
<tr>
<td></td>
<td>$\eta$</td>
<td>3983 &lt; 4240 &lt; 4513</td>
</tr>
</tbody>
</table>
V. UNCERTAINTY AND OPTIMIZATION

When making the simulations, two kinds of uncertainty must be taken into account: the one upon the model (confidence intervals of the parameters of the reliability laws) and the one inherent to simulation (standard deviation of the values of the indicators). Depending on the objectives and the acceptable risk we can consider, we can give bounds of the indicators associated with a percentage of risk. For instance, “80% of time, availability will be over 95%”. These conclusions appeared to be of the highest interest for the different maintenance managers we met during our studies made for industries. Indeed, there is often too much uncertainty around data to give exact conclusions.

In order to show the effect of the uncertainties, we will apply our methodology on a virtual system for which the set of operational data will be simulated by using fixed reliability laws (representative of our case study). We assume that the industrial CMMS allow a classification by failure mode. The first one is tied to “random” failures (electrical interruptions, random problems with raw material etc.) and the second one to the mechanical failures. For our example, we will compare two cases: in the first one the operational data are only failure ones (50 for each component or failure mode). In the second one, data for the second component are censored because of preventive maintenance. These data will be simulated by Monte Carlo simulations using the following laws: exponential with $\lambda=0.01$ (first component) and a Weibull law with $\beta=2$ and $\eta=100$. We will identify reliability laws in the two cases (by the maximum likelihood estimation) then simulate the behaviour of the system for different preventive maintenance assuming the parameters of the laws are either the identified ones or the bounds of the confidence intervals (CI).

In the case “no censoring”, we assume that we obtain the following reliability laws:

- Exponential $\lambda=0.0117$ with CI [0.085; 0.0150].
- Weibull $\beta=1.84$ ([1.46; 2.32]) - $\eta=108$ ([92; 126]).

When the second component is under a preventive maintenance made every 90 time units, the parameters for the Weibull law become $\beta=1.35$ ([0.83; 2.20]) - $\eta=127$ ([85; 191]). In this case, we can already conclude that the “worst case” (lowest limits of the CI), a preventive maintenance will be less effective than a corrective one because $\beta<1$, which illustrate the high bias due to preventive maintenance, even when the original $\beta$ was 2.

The simulations (50 per case) were made for a preventive periodicity from 20 to 200 t.u. The corrective maintenance costs for the two components and the preventive ones for the second component are those of the table 3. The first identification case (no censoring) produced the Fig.3 to 8, where the mean curve and the “95%” ones are drawn. There are no clear optima on these graphs. The conclusions we can give are the best periodicity interval that combine good availability, low costs and low standard deviation. We can indeed see that the highest the preventive periodicity is, the higher this deviation tends to be (or at least more “random”). When the maintenance managers have these results, they can choose the periodicity that best fit their production planning and availability and costs objectives. Fig.9 to 14 shows the results for the censored case. The effect of censored data on the identification (as smallest $\beta$) is quite visible: the maintenance costs increase and the availability is smallest for the same periodicity.

VI. CONCLUSIONS

Maintenance nowadays needs to become intelligent. Every available data source is needed to optimize its activities. We have exposed a framework that aims at being adapted to the largest panel of industrial cases and still robust. It relies on reliability block diagrams that concentrate different level of description of the systems. These models are then used to simulate the behaviour of the systems. Thanks to the estimation of the performances of the system, maintenance managers can build their strategy knowing the risk not to achieve their objectives, taking into account the uncertainty upon the model. Monte Carlo simulations are useful to quantify the dispersion of performances indicators, and the 95% confidence intervals allow guaranteeing optimal costs and availability for an appropriate choice of preventive maintenance periodicity.

Further researches aims at estimating the effect of no production. In the context of the economical crisis, some production lines must stop up to 50% of calendar time. The effect of theses stops must be integrated in our methods to make the maintenance even more efficient.

<table>
<thead>
<tr>
<th>TABLE III. MAINTENANCE COSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Preventive maintenance</td>
</tr>
<tr>
<td>Corrective maintenance</td>
</tr>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>fixed 500€</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENT

This research is part of the MINT project, which is a project from the “Marshall Plan”. We want to thank the Walloon Region of Belgium for its interest and its funding.

REFERENCES

Figure 9. Maintenance costs vs periodicity (mean case – censored data)

Figure 10. Availability vs periodicity (mean case – censored data)

Figure 11. Maintenance costs vs periodicity (worst case – censored data)

Figure 12. Availability vs periodicity (worst case – censored data)

Figure 13. Maintenance costs vs periodicity (best case – censored data)

Figure 14. Availability vs periodicity (best case – censored data)