



HAL
open science

Failure prognostic by using dynamic Bayesian Networks.

Kamal Medjaher, Jean-Yves Moya, Nouredine Zerhouni

► **To cite this version:**

Kamal Medjaher, Jean-Yves Moya, Nouredine Zerhouni. Failure prognostic by using dynamic Bayesian Networks.. 2nd IFAC Workshop on Dependable Control of Discrete Systems, DCDS'09., Jun 2009, Bari, Italy. pp.291-296. hal-00402938

HAL Id: hal-00402938

<https://hal.science/hal-00402938>

Submitted on 8 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Failure prognostic by using Dynamic Bayesian Networks

K. Medjaher, J.-Y. Moya, N. Zerhouni

FEMTO-ST Institute, UMR CNRS 6174 - UFC / ENSMM / UTBM
Automatic Control and Micro-Mechatronic Systems Department
24, rue Alain Savary, 25000 Besançon, France
(e-mail: kamal.medjaher@ens2m.fr)

Abstract: this paper presents a procedure for failure prognostic by using Dynamic Bayesian Networks (DBNs). The graphical representation of this tool is particularly well suitable for modeling complex systems, with non homogeneous sources of data and knowledge. Moreover, DBNs allow to deal with uncertainty which is an inherent property to any failure prognostic work, especially regarding the estimation of the Remaining Useful Life (RUL) before a failure. The DBN model can be also used to observe the propagation of the effect of any state of the model on the other remaining states. The proposed procedure is applied on a small industrial system to show its feasibility.

Keywords: Dynamic Bayesian Networks, Failure, Prognostic, Remaining Useful Life.

1. INTRODUCTION

Nowadays, requirements of quality, productivity, security and ecology push industrials to invest in more and more complex production systems. With age, these systems become vulnerable to failures and their maintenance expensive. To keep them in good operational conditions, maintenance activity should be considered as an integral part of the company's whole strategy. Maintenance activity aims at increasing availability of the production tool, reducing the costs, improving security and limiting the consequences of accidents on the environment. It can be corrective, systematic preventive, conditional or proactive (Muller et al. (2008)). Each one of these maintenance strategies has its advantages and its drawbacks. However, in order to be more reactive, proactive maintenance proves to be the best maintenance to adopt. This is done by estimating the future health state of the machine by taking into account its actual state and the future conditions of its exploitation in order to better plan the maintenance actions. One of the key features of proactive maintenance is industrial failure prognostic.

According to the ISO standard (ISO, 13381-1 (2004)), failure prognostic corresponds to the estimation of the operating time before failure and to the risk of existence or future appearance of one or more failure modes. This operating time before failure is well known as Remaining Useful Life (RUL). Three main prognostic approaches are proposed in the literature (Vachtsevanos et al. (2006); Lebold and Thurston (2001)): model based prognostic, data-driven prognostic and experience based prognostic. The first approach requires a mathematical model of the degradation to estimate the RUL. The second approach uses data provided by the monitoring system to predict the degradation. Tools and techniques employed in this approach are generally those used by the artificial intelligence community. The third approach proposes an estimation of

the RUL by using reliability models obtained from the historical data of the machine.

The present contribution falls within the second approach and uses Dynamic Bayesian Networks (DBNs) (Murphy (2002)). A Bayesian network (BN) (Pearl (1988)) is a directed acyclic graph, where the nodes represent random variables and the links the causal and dependence relations between the variables. A DBN is a mean of extending BN to represent the time evolution of the random variables. It can be used as a graphical model to represent causal stochastic systems. This graphical tool is well suited to model complex systems because it allows to represent in the same model non homogeneous knowledge, to take into account uncertainty and to predict the system's future health, which can be used to estimate the system's RUL. The present paper proposes a procedure for failure prognostic on relatively small industrial systems. The second section is dedicated to some definitions and terminologies used in the framework of prognostic, with an introduction to the main prognostic approaches. The third section deals with the paper's main contribution, followed by the application of this latter on a real system. Finally, some concluding remarks are given in section five.

2. FAILURE PROGNOSTIC FRAMEWORK

2.1 Definitions and terminologies

The term prognostic finds its origin in the Greek word "*progignôskein*" which means "to know in advance". Regarding industrial prognostic, many definitions are given in the literature (see Muller et al. (2008); W.Q. Wang et al. (2004); Byington et al. (2002); Lebold and Thurston (2001) for more details). However, in this paper, only that one proposed by the standard (ISO, 13381-1 (2004)) is retained, and where prognostic is defined as the estimation of the operating time before failure and the risk of future

existence or appearance of one or several failure modes. This standard defines the outlines of prognostic, identifies the data needed to perform prognostic and sets the alarm thresholds and the limits of system's reset.

The first step of the prognostic process consists in monitoring the system. The data gathered are then pre-processed in order to be used by a diagnostic module to identify the actual operating mode. This state is then projected in the future in order to predict the system's future state. The intersection point between the value of each projected parameter or feature and its corresponding alarm threshold leads to what is known as RUL of the system (Fig. 1). As in any prediction work, a prediction error should

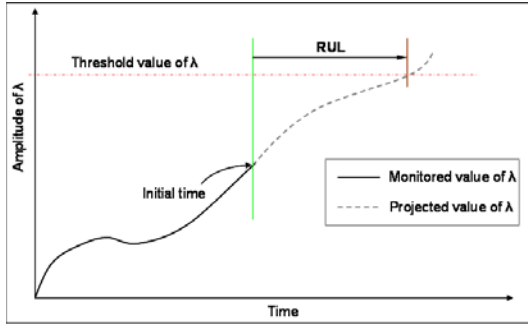


Fig. 1. The RUL's graphical representation

be associated to the estimated value of the RUL (Fig. 2). The sources of the prediction error may be multiple: modeling hypotheses, non-significant data, used prediction tool, uncertainty in the thresholds' values, etc. In addition, uncertainty is inherent to any prognostic (Provan (2003)). The error associated to any RUL estimation should de-

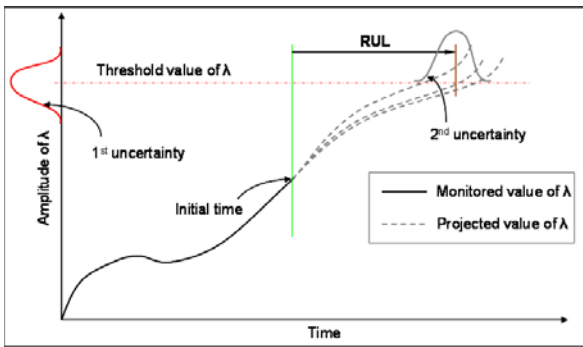


Fig. 2. Uncertainty associated to RUL estimation

crease as the time of the real failure approaches. This is exactly what happens in the case of weather forecast: the predictions given at the beginning of a week for the next Sunday, for example, are less precise than those given for the same day (next Sunday) but at one or two days before. This is because the predictions are adjusted each time new data are acquired.

Similarly to weather forecast, a confidence degree should be associated with any industrial prognostic work to render its conclusions more credible. Indeed, instead of telling an industrial that his/her machine will fail in x unites of time, it would be more realistic to give an estimated RUL with a confidence value. By including the uncertainty and confidence degree, the prognostic steps become more detailed as shown in Fig. 3. As mentioned previously,

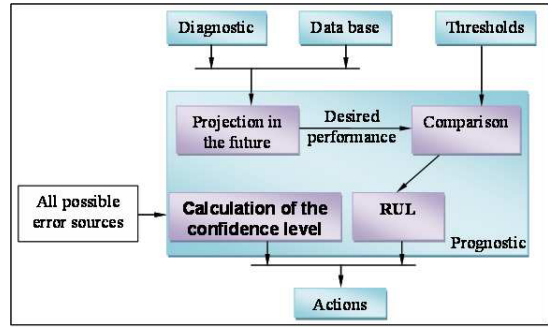


Fig. 3. Prognostic and confidence degree

the value of the estimated RUL is the output of some comparison between the projected state of the system and the predetermined threshold values. Note that, at the projection step, what is needed is not necessarily a value of a physical parameter but can be a desired performance, an achieved function or an availability of a service; depending on the kind of system on which prognostic is performed. After having given some definitions and terminologies used in the prognostic framework, the following section deals with the existing approaches, methods and techniques allowing to quantify the indicators previously introduced.

2.2 Prognostic main approaches

In the literature, there exists three main prognostic approaches summarized by Fig. 4 (Vachtsevanos et al. (2006); Lebold and Thurston (2001)). A survey of the methods used in each approach can be found in (Heng et al. (2008); A.K.S. Jardine et al. (2006)) and in the second chapter of Muller's thesis (Muller et al. (2008)).

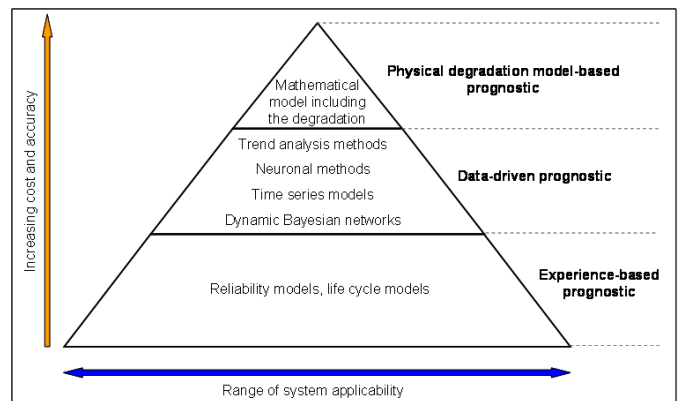


Fig. 4. Prognostic main approaches

Experience-based prognostic: it consists in using probabilist or stochastic models of the degradation, or of the life cycle of the components, by taking into account the data and knowledge accumulated by experience during the whole exploitation period of the industrial system. The probabilist model can be a simple probability function (Weibull, exponential, normal and Poisson laws) or a modeling in the form of stochastic process (Markovian or semi-Markovian).

The advantage of this approach is that it is not necessary to have complex mathematical models to perform prognostic. Moreover, this approach is easy to apply on systems

for which significant data are stored in a same standard that facilitates their use. For example, a company which has conserved during a long period of time a production and maintenance database with some minor rules and standards for data storing, can easily get the estimation of the parameters of the probability laws. However, the main drawback of this approach dwells in the amount of data needed to estimate the parameters of the used laws. Consequently, this approach can not be applied in the case of new systems for which data from experience feedback do not exist. The other kind of problem is that in most of cases, it is necessary to filter and pre-process the data to extract the useful features.

Data-driven prognostic: the principle of this approach consists in collecting information and data from the system and projecting them in order to predict the future evolution of some parameters, descriptors or features, and thus, predict the possible probable faults. Without being exhaustive, mathematical tools used in this approach are mainly those used by the artificial intelligence community, namely: temporal prediction series, trend analysis techniques, neuronal networks under all their facets, neuro-fuzzy systems, hidden Markov models and dynamic Bayesian networks.

The advantage of this approach is that, for a well monitored system, it is possible to predict the future evolution of a degradation without any need of prior mathematical model of the degradation. However, the results obtained by this approach suffer from precision, and are sometimes considered as local ones (for the case of neural networks and neuro-fuzzy methods).

The difference between experience-based prognostic and data-driven prognostic is that in the second approach most of the data are provided by sensors, whereas in the first approach the data are mainly those taken from experience feedback (number of breakdowns, replaced and repaired components, etc.).

Model-based prognostic: this consists in studying each component or sub-system in order to establish for each one of them a mathematical model including the degradation. The derived model is then used to predict the future evolution of the degradation (Luo et al. (2003); Chelidze et al. (2002)). In this case, the prognostic consists in evolving the degradation model till a determined future instant from the actual deterioration state and by considering the future use conditions of the corresponding component.

The main advantage of this approach dwells in the precision of the obtained results, as the predictions are achieved based on a mathematical model of the degradation. However, the derived model is specific to a particular kind of component or sub-system, and thus, can not be generalized to all the system. In addition to that, getting a mathematical model of a degradation is not an easy task and needs well instrumented test-benches which can be expensive.

3. FAILURE PROGNOSTIC BY USING DBN

This section presents a DBN-based procedure for failure prognostic. The choice of this tool among other modeling tools (like stochastic Petri nets, neuronal networks and classical time-series models) can be justified by the fact that a DBN allows to consider in the same and unique

graphical model different and non-homogeneous kinds of knowledge related to the system (global rather than local view), to take into account uncertainty and causality, and can be applied for both discrete and continuous cases.

3.1 Brief introduction of DBNs

A DBN is a mean of extending BN to represent the time evolution of the random variables. They generalize most of the stochastic models like Hidden Markov Models (HMM), Hierarchical HMMs and Kalman filter (Murphy (2002)). A DBN is an unrolled BN, which is graphically represented by two time slices (at times t and $t + 1$). Each time slice contains a static BN which remains the same. The dynamic aspect is not related to time evolution of the structure of the model, but to the time persistence of some phenomena. This persistence is represented by links between nodes of the slice at time (t) to the slice at time ($t + 1$). In a DBN, the Markovian hypothesis is respected. This means that the probability distribution of a node at ($t + 1$) depends only on the probability of its parents at (t) and on those in the same time slice, and not on those at ($t - 1$).

To build a DBN, one can proceed by two steps: the qualitative step, which consists in obtaining the graphical structure of the model, and the quantitative step for estimating the conditional probabilities of the random variables. Several available algorithms allow to learn DBN's structure and to estimate the CPTs of the derived model (Murphy (2002)). Once the DBN obtained, it can be used in filtering, smoothing or prediction. In the present work, only prediction use, allowing to estimate the system's future health and thus the RUL, is considered.

3.2 Procedure for failure prognostic by DBNs

The following procedure is applicable on machines and components subjects to failures and faults, needing maintenance and for which one has a multitude of data and information provided by sensors, operators, tests, etc. Thus, it is assumed that the available knowledge is enough sufficient to build the graphical structure of the system's DBN. The main points of the procedure are summarized hereafter:

- **Data collection:** this concerns the historical data of the machine (report of failures and faults, achieved maintenance, etc.), the data related to the components (identify the descriptors of each component like temperature, pressure, the time persistent phenomena like wearing, fatigue, etc.) and the data related to the environment in which the machine is used.
- **Variables identification:** this consists in assigning a name, a descriptor and a measurable (or verifiable) degradation (or failure indicator) to each component of the machine. Also, the type (continue or discrete) of each descriptor must be chosen because this determines the type of data (numerical value, deterministic or logical law, etc.) of the corresponding Conditional Probability Table (CPT) of the descriptor.
- **Causal relations identification:** two types of causal relations can be considered, namely: static and dynamic relations. The first ones can be obtained from the knowledge one has about the interactions between the variables represented by nodes on each time slice

of the DBN. They can be given by an expert of the machine, deduced from a FMECA study or from the physical comprehension of the interactions between the physical components. The second ones, linking two or more variables, are derived from the comprehension of the degradation phenomena. These phenomena are persistent and thus, their corresponding causal relations are represented by links between two time slices of the DBN. Of course, this involves a well established degradation models allowing to deduce the numerical values of the corresponding CPT.

- Construction of the DBN: once the variables and their static and temporal dependences are established, the graphical structure of the DBN can be constructed. As for any modeling methodology, the precision and the quality of the obtained DBN depends on the modeling hypotheses. In addition, one can note that the nodes of the DBN do not represent necessarily physical phenomena, but can represent any kind of event.
- Definition of the CPTs: three type of CPTs are defined in this procedure. The CPTs of the “statistical” nodes (obtained by statistical calculations from an observation database), the CPT of the “expertise” nodes (the numerical values are given by an expert of the machine according to a probability scale) and the CPTs of the “dynamic” nodes (can be obtained from degradation mechanism laws, by a statistical calculation or by expertise).
- Definition of the thresholds: this task is important and is not trivial to achieve. This is because the precision of the prognostic results and, consequently, the maintenance actions which will be planned are directly related to the threshold values. These latter should be chosen so that to avoid false alarms (value not very low), and to have sufficient time between the moment when the alert is given and the moment of maintenance intervention (value not very high).
- Placement of additional sensors: if the monitoring system is not sufficient to estimate the CPTs of the DBN, additional sensors may be installed to remedy to this situation.
- Prediction, evaluation and comparison: in this step, available inference algorithms are launched in order to calculate the probability distribution of any variable of the DBN, simulate the future state of the machine and estimate the numerical value of the RUL.

4. APPLICATION TO THE SISTRE PLATFORM

4.1 System description

The SISTRE platform is a small production system used by the AS2M department for its research activities and for technologies transfer in the field of industrial automation. It is also an important support for the department’s e-maintenance platform, and for the teaching activities. The system includes common functionalities such as displacement, manipulation, assembling and takes benefits from the most recent technology innovations like wireless remote control via WIFI. The platform is composed of five stations almost identical, one principal inner conveyor and one secondary external conveyor. Each station is controlled by

its own automaton and the data are stored in a dedicated computer. The operating principle of the platform is easy.

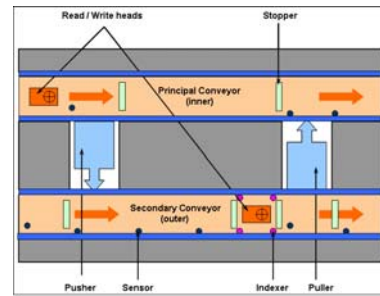


Fig. 5. A typical station scheme

The conveyors push the pallets which initially are on the principal conveyor. To move from one station to another station for a particular treatment, the pallets are pushed toward the secondary conveyor. Each pallet is equipped by an “electronic label” in which is recorded the list of operations the pallet will undergo at each station. Before each station, a reader-writer module reads the list of operations and decides whether or not the pallet will go through this station. This module can also modify the data initially recorded on the label (for example: “the given operation is successfully done” or “that sensor is ON at that time”). The main components of a typical station are (Fig. 5):

- 1 pusher: pneumatic jack which decants the pallets from the main conveyor to the secondary conveyor,
- 1 puller: pneumatic jack which decants the pallets from the secondary conveyor to the main conveyor,
- 1 indexer: pneumatic jack allowing to lift the pallets from the conveyor, insuring at the same time a correct positioning for the operation the pallets will undergo,
- 2 conveyors (one inner and one outer): composed of 2x2 strips carried by a synchronous electrical motor via a chain-shaft pulley system,
- 8 to 9 stoppers: pneumatic jacks used to stop the pallets by opposition, the strips slipping always under the pallets,
- 8 to 9 inductive sensors: used to indicate the presence or absence of a pallet,
- 2 read/write head of BALOGH type: used to read the data stored in the label of each pallet, and to write the account of the achieved operation.

4.2 DBN of a standard station

The study of a typical station has shown that most of the encountered faults are due to problems of the belt (positioning, blockage, wearing and lengthening) and to the failure of the electrical asynchronous motor. The functional and dysfunctional analysis performed on a typical station in addition to the discussions with the expert of the platform and the application of steps two to four of the prognostic procedure described in sub-section 3.2 have led to the DBN of Fig. 6. In this DBN, the node “Operational” is used to model the availability or not of the pallets’ routing movement, which depends on the state of the physical components of the station, represented by the rest of the nodes in each time slice. The causal relations between the DBN’s variables are obtained by analyzing the mechanical operating of the pallets’ routing movement

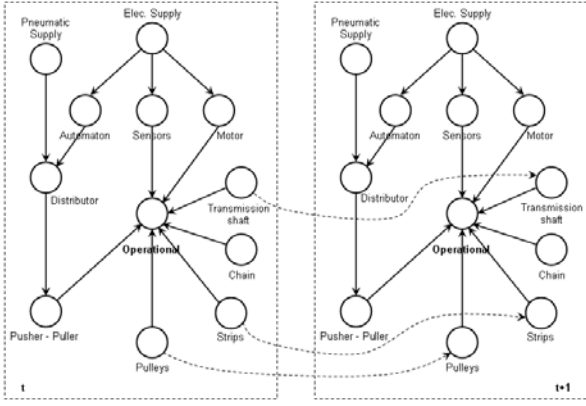


Fig. 6. A DBN of a typical station

Table 1. The CPT of the pulleys

Pulleys at t	Pulleys at $t + 1$		
	Ok	Degradation	Out of Service
Ok	99.92	0.07	0.01
Degradation	0	99	1
Out of Service	0	0	100

(for the static nodes in each time slice), and degradation mechanisms for the persistent nodes (links between nodes of two different time slices). For example, the node “Strips” is a persistent node representing the wearing phenomenon of the strips. The state of this node at $(t + 1)$ depends only on the state of the same node at (t) . The CPT of the nodes in the DBN of Fig. 6 are obtained by three ways: boolean modes for the node “Operational”, reliability functions for the nodes representing the degradation mechanisms, and statistically or from the expert’s knowledge for the rest of the nodes. As an example, for the pulleys, the wearing of their ball-bearings is taken into account in this study. The corresponding CPT of the node “Pulleys” is then determined from a reliability function given by the manufacturer of these ball-bearings:

$$F = \exp \left(\ln 0.9 \left(\frac{L}{L_{10}} \right)^\beta \right), \quad (1)$$

where $\beta = 1.5$ is the slope of the Weibull associated function, $L_{10} = \left(\frac{C}{P} \right)^n \frac{10^6}{60 \cdot V}$ is a time duration in hours with $C = 2500 \text{ N}$, $P \approx 51 \text{ N}$, $n = 3$ and N is the rotating speed in solicitation. By considering a time unit of $1h$, the numerical value of the probability that the pulley fails is equal to $1 - F = 10^{-12}$. This value is tiny because the ball-bearings of the pulleys are in fact over-dimensioned in the platform (this was confirmed by the expert of the platform). To show the feasibility of the method, in the coming simulations the value of the failure probability is taken more higher than that one obtained by the previous calculations. In this study, the used CPT of the node “Pulleys” is given by Table 1. The definition of the thresholds, allowing to estimate the numerical value of the RUL, is based on the availability of the function “Operational” at 50%.

4.3 Simulation results

The DBN of a typical station is simulated by using BayesiaLab software (Bayesia (2008)). The time unit used

in this simulation is 10 hours. The DBN implemented in this software is given in Fig. 7, where the temporal dependencies are indicated by red links. The Simulation

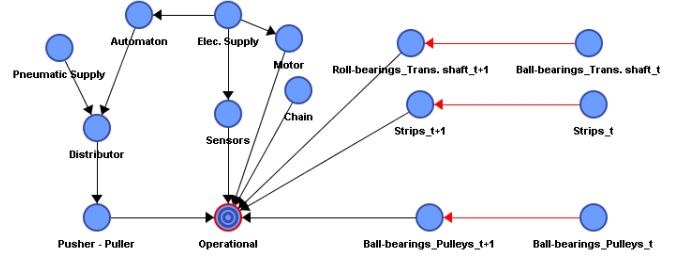


Fig. 7. The DBN for failure prognostic

results obtained when introducing all the degradation mechanisms of different components of the station are given in Fig. 8. It can be observed that the probability of

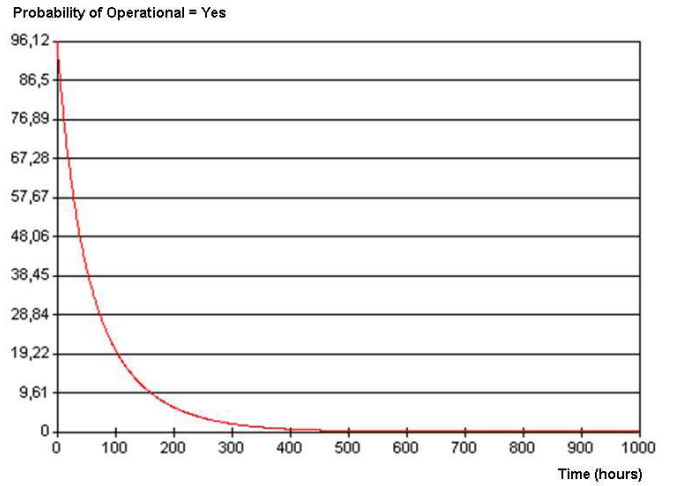


Fig. 8. The time evolution of the node “Operational”

the function “Operational” reaches 50% at about 36 units of time (which gives the RUL’s numerical value). This value is in fact low due to the contribution of all the degradation mechanisms and to the fact that no maintenance action is introduced, which accelerates the unavailability of the principal function. However, when introducing maintenance actions (strips replacement at 36 units of time interval), the availability of the pallets’ routing movement increases as shown in Fig. 9. It can be observed that the availability of the function “Operational” goes up each time the strips are replaced. However, the global form of the curve tends to decrease below 50% of availability. This observation is correct and is due to the effect of the ball-bearings wear that is not maintained. To remedy to this situation, a maintenance action of the pulleys’ ball-bearings is introduced each 1100 units of time, which leads to the results shown in Fig. 10. It can be seen that if the ball-bearings are not replaced, the curve has tendency to go down below 50% of availability of the function “Operational” in spite of regular replacement of the strips. But, as the ball-bearings are replaced, the curve goes up. However, this probability does not reach 100% because of the influence of the ball-bearings of the transmission shaft which are not maintained. Note that the defined threshold at 50% of availability of the function “Operational” is not

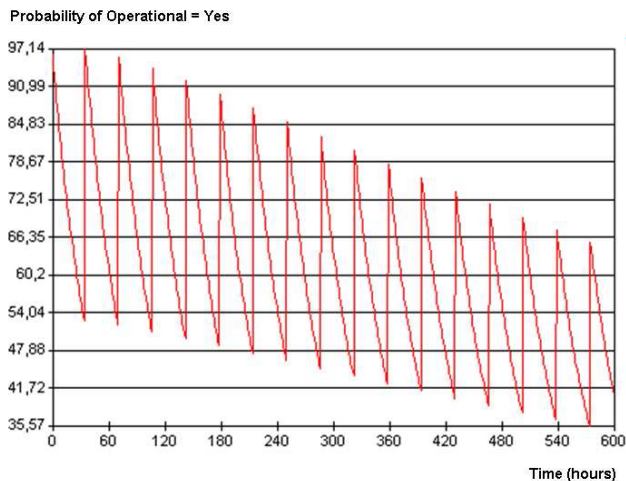


Fig. 9. Introduction of a maintenance action

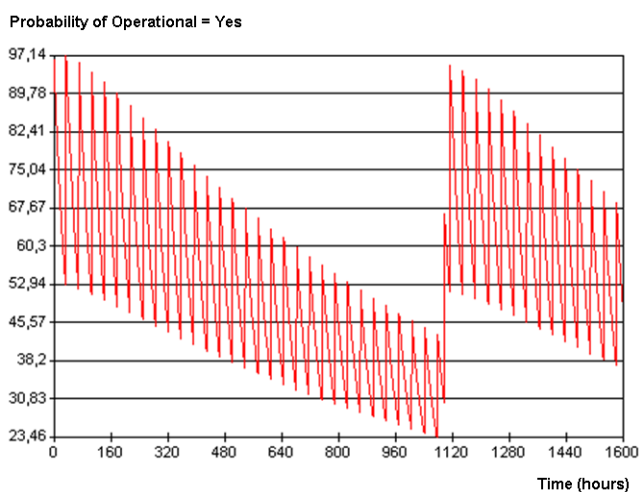


Fig. 10. Replacement of the ball-bearings

respected. To do so, the ball-bearings would be replaced each 210 units of time (see Fig. 10), but this would mean to change them while they are still in a good state.

5. CONCLUSION

This paper has shown the applicability of DBN in failure prognostic on a small real manufacturing system. The graphical structure of the DBN was built based on the available knowledge about the system's behavior, the degradation mechanisms, the functional decomposition and the links between the system's components. Once this structure derived, its quantification (the estimation of the CPT of each node of the graph) has been achieved, which led to the final exploitable DBN for failure prognostic. Contrary to fault prognostic by methods like those using physical or analytical (mathematical) models, DBN-based fault prognostic allows to estimate the value of the RUL by using different indications (availability of a function, service, component, etc.). In addition, the graphical representation of DBN can be exploited to propagate the state of any function or component, represented by its corresponding node, and observe its influence on the states of the remaining nodes of the graph. The estimated value of the RUL can be used to plan new maintenance actions,

and thus allow reducing the part of the systematic preventive maintenance and increasing that one of the proactive maintenance with all the benefits this latter offers in industrial applications.

However, in this contribution, the numerical values of the CPTs were obtained either by experience or from reliability laws. The estimated value of the RUL can be improved by estimating on line the real values of the DBN by taken into account the data provided by the sensors installed on the platform. This is in fact a big challenge on which our future works are oriented.

REFERENCES

- Bayesia (2008). Bayesialab software. www.bayesia.com.
- Byington, C., Roemer, M., Kacprzyński, G., and Galie, T. (2002). Prognostic enhancements to diagnostic systems for improved condition-based maintenance. In *2002 IEEE Aerospace Conf., Big Sky, USA*.
- Chelidze, D., J.P. Cusumano, and Charterjee, A. (2002). A dynamical systems approach to damage evolution tracking, part 1: The experimental method. *Jour. of Vib. and Acoustics*, 124, 250–257.
- Heng, A., Zhang, S., A.C.C. Tan, and Mathew, J. (2008). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*. In press.
- Lebold, M. and Thurston, M. (2001). Open standards for condition-based maintenance and prognostic systems. 5th annual maintenance and reliability conference (marcon 2001). Gatlinburg, USA.
- Luo, J., Namburu, M., Pattipati, K., Qiao, L., Kawamoto, M., and Chigusa, S. (2003). Model-based prognostic techniques. In *Proc. of IEEE AUTOTESTCON 2003*, 330–340.
- Muller, A., M.C. Suhner, and Iung, B. (2008). Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system. *Reliability Engineering and System Safety*, 93, 234–253.
- Murphy, P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. Ph.D. thesis, Univ. of California, Berkeley.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufman Publishers.
- Provan, G. (2003). Prognosis and condition-based monitoring: an open systems architecture. In *5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, 57–62.
- A.K.S. Jardine, Lin, D., and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. and Sign. Proc.*, 20, 1483–1510.
- ISO, 13381-1 (2004). *Condition monitoring and diagnostics of machines - prognostics - Part1: General guidelines*. Int. Standard, ISO.
- W.Q. Wang, M.F. Goldnaraghi, and Ismail, F. (2004). Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, 18, 813–831.
- Vachtsevanos, G., F.L. Lewis, Roemer, M., Hess, A., and Wu, B. (2006). *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. New Jersey, Hoboken: Wiley & Sons.