System approach-based Bayesian Network to aid maintenance of manufacturing process

Philippe Weber, Marie-Christine Suhner, Benoît Iung

SYSTEM APPROACH-BASED BAYESIAN NETWORK TO AID MAINTENANCE OF MANUFACTURING PROCESS

Weber P., Suhner M.-C., Iung B.

Centre de Recherche en Automatique de Nancy (CRAN), CNRS UPRESA 7039
Université Henri Poincaré, Nancy 1
2, rue Jean Lamour
F-54519 VANDOEUVRE-LES-NANCY Cedex – FRANCE
Email: [philippe.weber,suhner@esstin.uhp-nancy.fr,benoit.iung@cran.uhp-nancy.fr]

Abstract: The prospective work reported here explores a new methodology to develop Bayesian Network-based diagnosis and prognosis aids for manufacturing processes. This work is justified with the complex systems by the need of controlling and maintaining in dynamical way the global system performances in order to optimise the enterprise strategies. The added value of our methodology is to formalise the maintenance aid models from a priori knowledge both on the system functioning and malfunctioning by means of Bayesian Networks. The networks are built on adaptability principles and integrate uncertainties on the relationships between causes and effects. The feasibility of this methodology is tested in a manufacturing context with the maintenance aids on a lathe machine.

Keywords: maintenance, Bayesian Network, system approach, diagnosis strategy, prognosis strategy, FMECA.

1. INTRODUCTION

One of the main challenges of the Extended Enterprise is to maintain and optimise, in dynamics, the quality of the services delivered by industrial objects all along their life cycle. These objects are increasingly complex and constrained today by societal, economical and technological environments having more and more requirements. During the object-manufacturing phase, the improvement of the global performances of the manufacturing system is an issue to be solved to ensure the optimisation of the Enterprise needs. In that way, to control the direct and indirect costs of the system compared with the necessary availability, is a first challenge raised.

The objective is thus to have maintenance processes mainly in terms of decision-making aids in order to guarantee maximum components availability keeping the system in operation. These processes do not have the aim to replace the operator but to assist him in its final decision-making to avoid consequences of an error in judgement.

Today, most of current automated systems do not provide the means for intelligent interpretation of information copying with large process disturbances (diagnosis) and for predicting the consequences of a future action (prognosis). However tools issued from Artificial Intelligence (Khoo, *et al*., 2000), (Hu, *et al.*, 2000) are used now to bring a first decision-making aid for manufacturing systems. In addition, recent works on system safety and Bayesian Networks (BNs) are developed (Kang and Golay, 1999). For example, Bouissou, *et al*. (1999) propose, within the SERENE project, a hierarchical decomposition of the decision-making model for system safety analysis. As for them, Bobbio, *et al.* (2001) explain how the Fault Tree can be achieved using BNs.

In relation to these works, the proposed methodology has originality on formalising, by means of BNs (section 2), the diagnosis and prognosis aid models (section 3) from a priori knowledge on the system.
So, issued from design, the networks are built:

- by inheriting the adaptability principles as advocated by the IMS (Intelligent Manufacturing Systems) initiative; it is a first answer to knowledge structuring and reusability as looked for in the expert systems,
- by incorporating uncertainties on the relationships between causes and effects to be more consistent with system reality.

To show the feasibility and the added value of this prospective methodology, an application is developed in section 4 on the case of maintenance aid for lathe machine. Finally, in section 5, conclusions and prospects are presented.

2. BAYESIAN NETWORK THEORY

The BNs are Directed Acyclic Graph (DAG) and used to represent uncertain knowledge in Artificial Intelligence (Jensen, 1996). A BN is defined as a pair: G=((n,e),CPT), where (n,e) represents the DAG; “n” is a set of nodes, defined by different states; “e” is a set of directed edges describing the probabilistic dependencies between nodes. In this work, the nodes represent discrete random variables of the process. In pair G, each node is associated to a Conditional Probability Table (CPT) to quantify the dependencies between random variables as conditional probabilities. A node without parent is called a root node.

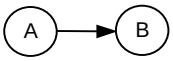

Fig. 1. Basic example of a BN.

A probability is allowed to each state of the node. This probability is defined, a priori, for a root node and computed by inference for the others. The computation is based on the probabilities of the parents states and the CPT. For instance, two nodes A and B with each of them two states (S_{*1} and S_{*2}) compose the BN in Fig. 1. The a priori probabilities of the node A are defined as:

A	S_{A1}	$P(A=S_{A1})$
	S_{A2}	$P(A=S_{A2})$

The probabilities of the states allowed to B are computed using a CPT. This CPT is defined by the probability of each B state knowing the state of A.

A		S_{A1}	S_{A2}
B	S_{B1}	$P(B=S_{B1}\|A=S_{A1})$	$P(B=S_{B1}\|A=S_{A2})$
	S_{B2}	$P(B=S_{B2}\|A=S_{A1})$	$P(B=S_{B2}\|A=S_{A2})$

Thus to compute $P(B=S_{B1})$, the following eq. 1 is used:

$$\begin{aligned} P(B=S_{B1}) &= P(B=S_{B1}|A=S_{A1}).P(A=S_{A1}) \\ &+ P(B=S_{B1}|A=S_{A2}).P(A=S_{A2}) \end{aligned} \quad (1)$$

The added value of a BN is linked to the computation of the probabilities allowed to a node state, knowing the state of one or several variables. If the knowledge on the process modelled by the BN is unavailable then the computation is based on a priori probabilities. However if the knowledge is increasing, the computation takes it into account and the results are adapted to the process state. It is then possible to estimate the impacts of a random variable on the process.

3. MODELLING APPROACH

The proposed modelling approach consists, from functioning systemic analysis based on SADT[1] graphical representation, (a) in representing the abnormal operation (malfunctioning) based on FMECA[2] and then (b) in formalising and unifying these two results in a unique model by means of BNs.

3.1. From system complementary functioning-malfunctioning representation…

The functioning and malfunctioning of the system are dual and must be studied together to control each system variable (Léger, *et al.*, 1999). It leads, first, to focus on the system functioning in relation to its environment and

[1] Structured Analysis and Design Technique
[2] Failure Mode, Effects and Criticality Analysis

its global internal and external resources. This action can be made by means of a **functioning modelling** using SADT graphical representation. This modelling is based on the principle of **activity** and sub-activities until elementary activities, supported by components, are emerging. Each activity (Fig. 2) fulfils a **finality**, which is to modify a "product" carried out by the manufacturing system. It produces or consumes flows such as "Having to Do" (HD) materialising the Input/Output (I/O) finality, "Knowing How to Do" (KHD) materialising the I/O knowledge, "being Able to Do" (AD) representing I/O energies, resources, activity support and finally "Wanting to Do" (WD) materialising the I/O triggers.

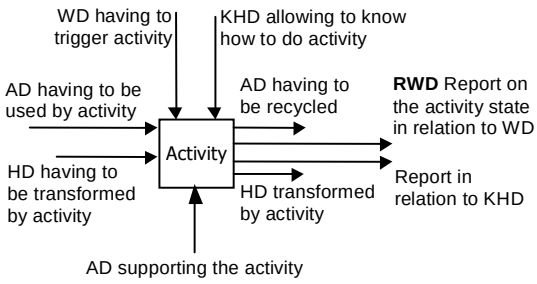

Fig. 2. Flows and Activity Representation.

For example, the output flow WD is a report (RWD) that represents the informational result of the Input HD product flow transformed by the activity.
From this functioning, the malfunctioning is induced (Léger, *et al.*, 1998) by considering that the relationship between these two modes is directly linked to the relationship between the normal and abnormal states of the system. The malfunctioning modelling can be made by means of FMEA study in order to identify the failure or degradation modes of each activity, the elements which are at the failure origin (causes) and the possible consequences of these failures (effects).
For example, the RWD flow can take the value "fulfilled or realised (R)" corresponding to the nominal state of the activity or the values "No true to the nominal or Not conform (NC)"; "Not fulfilled or Not Realised (NR)" requiring to identify the causes and the effects related to these two abnormal states.
The failure causes are either external with the activity (Input flows) or internal ones because linked to the AD activity support flow (components). A whole of states can be thus associated with each component. These states correspond to: nominal operation (OK), breakdown 1, breakdown 2… In the same way, the consequences are observable either on activity output flows or on the influence of the component degradation development on itself (to go towards a breakdown state). To sum up, a failure cause leads to a failure mode (e.g. the modification of the function state reported in RWD), which has as effect for the function not to produce nominal flow HD any more. This FMEA study is completed by a FMECA study where each failure mode is characterised by a criticality (risk priority number) resulting from the product of the three criteria, which are the Frequency (F), the Severity (S) and the No Detection (ND) (Suhner, *et al.*, 1992).

3.2. … to unified Bayesian Network representation

The BN is directly built from the dual functioning/malfunctioning analysis presented above leading to a unified representation. This representation is structured as a BNs tree. Its root is a BN representing the highest abstraction level. The elementary activities represent the lowest functional levels modelled by BNs.
To keep the generic function concepts (Fig. 2), its inputs are modelled, in the BN, by input nodes defining the random variables associated to the flows AD, HD, KHD and WD.
From this step, simplifying assumptions are made to carry out our prospective study until a feasibility phase. Therefore just the RWD flow is taken into account as main output. This flow is an informational view of the function finality, so it is assumed to be the same as the added value on product flow represented in the Fig. 2 by the HD flow transformed by the activity. It is thus transferred as informational view of physical result through the input flow HD of another function. The generic function represented in BN formalism is given in Fig. 3.
On this basis, the BNs are built in different ways according to the abstraction level and the design project advancement.

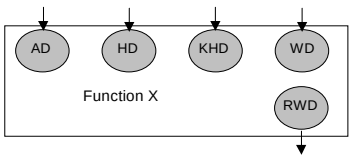

Fig. 3. Generic BN input and output nodes structure.

The high functional levels. To model high functional levels, the BNs are composed of generic sub-functions with

the structure defined in Fig. 3. The number of nodes states is adapted and it is possible to duplicate several inputs or outputs nodes (AD, HD…) if the function carries out several missions. Moreover, it is possible to model sub-functions in parallel or in series (Fig. 4).
In Fig. 4, as the generic sub-functions F1 and F2 are in series, the report RWD1 is transferred to F2 through the input flow HD. As the functions F2 and F3 are in parallel, a node OR is linked to RWD2 and RWD3 to compute the RWD of the global function. In this level, only the connections between functions in parallel are defined as CPT.

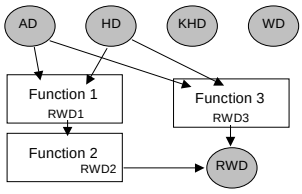

Fig. 4. High level of the functional decomposition.

The CPT of a OR node, for two functions whose RWD2 and RWD3 are composed of three states (R, NC, NR cf. section 3.1), is defined as follows:

RWD2	R			NC			NR		
RWD3	R	NC	NR	R	NC	NR	R	NC	NR
R	1	1	1	1	0	0	1	0	0
NC	0	0	0	0	1	1	0	1	0
NR	0	0	0	0	0	0	0	0	1

The low functional levels. The last BN of a branch is based on the FMECA analysis. The nodes constituting this BN are:

- the component nodes (CMP) or a set of components nodes according to the model explanation,
- and elementary function nodes (EF).

The CMP are defined only in the low BN levels. They are directly linked to the EF nodes representing their functionality. The CMPs states are defined by the causes analysed by means of FMECA. The causes are either internal of the low BN level i.e. linked to CMP, or external i.e. linked to the input nodes. The common causes are defined in higher hierarchical levels and the information forwards by heritage between the levels through the input and output nodes. There is thus duplication either for the CMPs or for the external causes. The EF nodes are linked to the CMPs and the input nodes leading to compute the RWD states probabilities (Fig. 5). If all the EFs are realised then the RWD is realised.

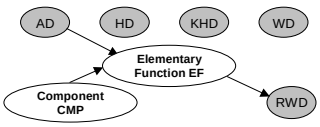

Fig. 5. Low level of the functional decomposition.

Evaluation of the probabilities. The FMECA analysis assigns a criterion F to the causes of failure. A probability of occurrence is associated with F. For sake of simplicity, only this criterion is presented. Nevertheless information such as the Severity (S) and the No Detection (ND) can also be considered. On this basis, the parameter F is used to determine the a priori probability that a CMP is in a state among a list of causes (considered as exhaustive). An example of F and causes is described below (with F = 2 when one failure occurred / month, and 3 for one / week):

Failure Modes	Causes	Effects	F
Elementary function in mode 1	CMP failure P1	Effect 1	2
	CMP failure P2	Effect 2	3

The a priori probabilities associated to the CMP node states in the BN are estimated for a given operating time (T), for example T=2600 h / year:

$$P(CMP = P1) = 0.005\ h^{-1} \qquad (2)$$

$$P(CMP = P2) = 0.020\ h^{-1} \qquad (3)$$

The a priori probability of correct operation of CMP is then deduced by:

$$P(CMP = OK) = 1 - (P(CMP = P1) + P(CMP = P2)) \qquad (4)$$

So, the a priori probabilities for each cause are:

- correct operation (OK): 0.975 h^{-1}
- failure 1 (P1): 0.005 h^{-1}
- failure 2 (P2): 0.020 h^{-1}

The EF states are defined by the failure modes. For instance, the EF nodes in the BN take the following states:

- Realised (R),
- Not Conform or partially realised (NC),
- Not Realised (NR).

No a priori probability is associated to these states because they are calculated according to the states of their parents i.e. the internal causes (CMP node states) and the external causes (input node states).
The CPT of the EF is defined thanks to the columns of the causes and the modes of the FMECA analysis.

AD		R			NC			NR
CMP		OK	P1	P2	OK	P1	P2	*
EF	R	1	0.2	0	0	0	0	0
	NC	0	0.8	0	1	0.2	0	0
	NR	0	0	1	0	0.8	1	1

The probabilities link the causes to the EF modes. Nevertheless the power of the BNs representation is that a combination of causes (for instance AD=R and CMP=P1) can lead to several failure modes with different probabilities allowing then to model the uncertainty.

A prognosis model. The BN model defined above allows the analysis of the degradation influences on the flows produced by the activity. This analysis is based on the simulation of a component failure, a common cause or an unconformity of a sub-function. The objective is to forecast the impacts of failures on the functions. It is then possible to analyse the upstream and downstream consequences on the whole system. For example, if a component failure is supposed, an evidence is defined as for instance P(CMP=P1) = 1. Then the probabilities associated to sub-functions are updated after the BN's inference. The RWD of each function relates the failure impacts on each functional level. Thus in a design phase, the BN model allows to aid the designer in his decision-making by the estimation of the various impacts of the solutions (Weber and Suhner, 2001).

3.3. Diagnosis

The diagnosis process tries to determine the origin of the degradations or the failures. Based on the observed situation, the causality represented as a relation between the activities leads to determine **internal cause** or **common cause**.

Principle. After an observation of a failure mode on the process, the evidences are defined in the BN. These modes correspond in states R, NC or NR of the observed variables. The BN's inference computes the probabilities of the unobserved variables states in order to determine the most probable causes. The cause with the most significant probability is suspected. This cause indicates the component, which is, a priori and according to the knowledge, the origin of the observed failure mode. It is then necessary to check its state. If the component is at the origin of the failure, then the diagnosis is finished. Otherwise, the knowledge is considered as not enough sufficient. In this case, the component state is taken into account as new evidence and another BN's inference determines a new cause.

Diagnosis aid with BN modelling. To facilitate the investigation into the most probable cause, the diagnosis is fulfilled using the hierarchical decomposition of the BNs model. So, a solution consists in inserting diagnosis nodes for each functional level to specify the **cause location** as: internal, upstream or downstream from the level. The "Input Diagnosis" node is connected to the input nodes of the function and defines if they are conforms to carry out the mission. The "Internal Diagnosis" node is connected to the RWD flows to evaluate if the mission is carried out (RWD) (Fig. 6). The states of these diagnosis nodes are: R, NC and NR.

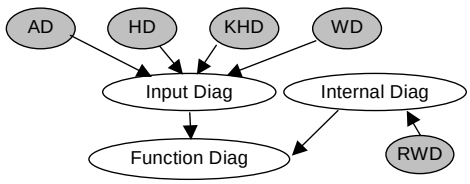

Fig. 6. Diagnosis aid nodes.

"Input Diagnosis" node and "Internal Diagnosis" node are connected to the "Function Diagnosis" node resulting in the cause location.
The rules allowing to define the CPT of the "Function Diagnosis" node are the following:

- The cause location is **internal** if the function is not correctly realised (Internal Diagnosis=NC or NR) and the input nodes are conform to carry out the mission (Input Diagnosis=R).
- The cause location is **upstream** if the input nodes are not conform to carry out the mission (Input Diagnosis=NC or NR), independently to the internal diagnostic.
- The cause location is **downstream** if the mission is carried out (Internal Diagnosis=R) and the input nodes are conform to carry out the mission (Input Diagnosis=R).

4. APPLICATION

The application chosen to test the feasibility of the approach is a lathe. Its functional analysis (activities and flows) of which the level A0 is presented in Fig.7, is composed of three main functions (to grip/to loosen the part, to put in rotation the part, to machine the part).

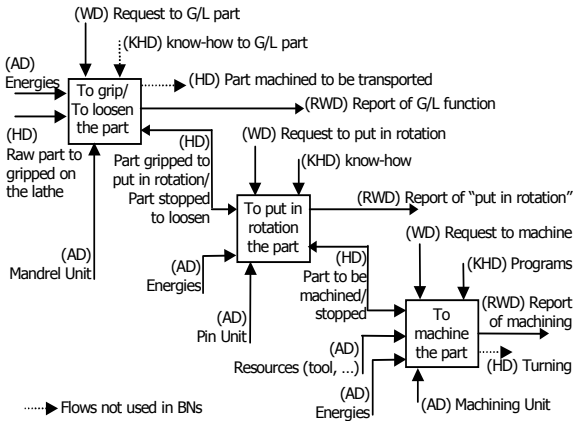

Fig. 7. Level 0 of the lathe function analysis.

Each function is broken down into one level at least allowing to make the link by a "physical support mechanism" with the real components of the lathe system. For example, the function "to grip/to loosen" is broken down into elementary generic functions such as "pre-actuate", "actuate". The function "pre-actuate" is supported by a distributor and the function "actuate" by a jack.

4.1. BN's Model

The software used to build BNs model is Serene:
(web link: http://www.hugin.dk/serene/).

Level 0: Function "Lathe" (Fig. 8). The BN represents high level of the functional decomposition of the lathe. Input nodes are grey: AD (electric and pneumatic power), WD (request to transform), KHD (programs), and HD (the raw part).

Table 1. Part of FMECA analysis on the function "actuate" supported by a jack.

Function	Element	Failure Modes	Effects	Causes	F	S	ND
To Actuate Gripping / Loosening	JACK	No functioning for "going in" moving	Jack shaft is not moving for "going in" (HD)	No energy to make "going in" action (WD) or Jack is Blocked (AD)	1	4	2
		No functioning for "going out" moving	Jack shaft is not moving for "going out" (HD)	No energy to make "going out" action (WD) or Jack is Blocked (AD)	1	4	2
		"Going out" action is too slow	Jack shaft is moving too slow for "going out action" (HD)	Not enough energy for "going out" action (WD) or Jack shaft is Jammed (AD)	3	2	2
		"Going in" action is too slow	Jack shaft is moving too slow for "going in action" (HD)	Not enough energy for "going in action" (WD) or Jack shaft is Jammed (AD)	3	2	2

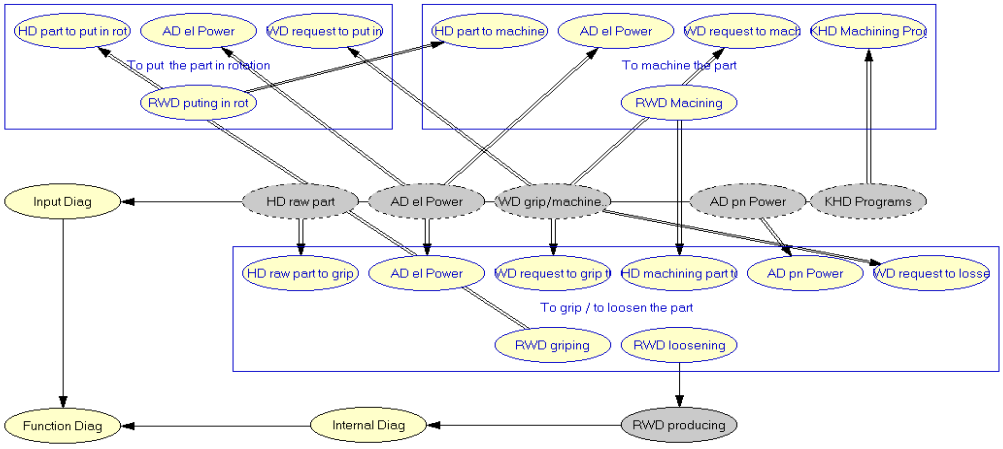

Fig. 8. Level 0: High BN.

The sub-functions are represented as sub BN (framed nodes), input nodes are linked to power, programs, requests, raw part and the output represents the RWD flow of the sub-function. The function "to grip/to loosen the part" is considered as composed of two sub-functions. The inputs and the outputs of this function correspond to gripping or loosening according to the state of the part. There is duplication of the input nodes HD and output nodes RWD, nevertheless any cycle is made in the networks thanks to the tree decomposition adopted.

Level 012: Function "Actuate" (Fig. 9). The BN presents the low functional level given by the FMECA analysis of the sub-function "actuate". The Fig. 9 represents the component "JACK" and its different EF: "going in" and "going out". These EF are linked to the "pneumatic power" and the "JACK". The "Internal and Input diagnosis nodes" are linked to the "Function Diagnosis" node.

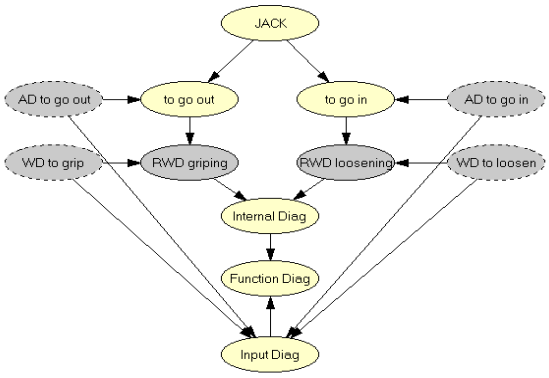

Fig. 9. Level 012: Actuate BN.

First inference. The a priori probabilities associated to the states of the variables representing the parts are calculated with an inference of the BN. The HD raw parts and RWD of each sub-function are presented in Table 2. Thus according to the characteristics of reliability used for the components, and as output of this lathe, 65.5% of the parts are machined correctly, 22.7% of the parts are machined but with no conformities, and 11.7% are not machined. This calculation is carried out according to the quality of the raw parts. Then the raw parts machined in conformity knowing the raw parts in state R before machining, are 72.8%.

4.2. BN inference Diagnosis

The diagnosis starts when the "RWD producing" is not in conformity or not realised. Let us develop the diagnosis whose scenario is the following: on the level 0 of the general function, the RWD producing is not in conformity NC.

Initially, input flows (power electric, pneumatic, and raw parts) are checked. The checking leads to consider these flows as all in conformity (state R).
The Function Diagnosis nodes of each sub-function of the level 0 are examined after the BN inference:

Diagnosis	To grip/ To loosen	To put in rotation	To machine
Internal	0.6438	0	**0.7122**
Upstream	0.3562	0	0
Downstream	0	1	0.2878

The function suspected is a priori "To machine the parts" because the probability of "Internal diagnosis" is the strongest. Thus the low level of this function is checked. The states of the variables representing the elements implied are observed and taken the following states:

LATHE	
OK	1
HS	0

TOOL	
OK	0.2878
Wear	0.7122
Broken	0

It is a priori "TOOL" abnormal wear which is the cause of the no conformity of the parts. The diagnosis is coherent; the LATHE is not incremented because a failure of the LATHE leads to parts in the state NR. Let us suppose that the checking of the tool shows that it is OK. Then the evidence is added to the knowledge:

TOOL	
OK	**1**
Wear	**0**
Broken	**0**

The states of BN are updated and a new inference proposes a new diagnosis taking account the new information. The procedure thus consists in going at the hierarchical level higher (i.e. the level 0). In this case, the failure mode is caused only from the sub-function "To grip/To loosen":

Diagnosis	To grip/ To loosen	To put in rotation	To machine
Internal	1	0	0
Upstream	0	1	1
Downstream	0	0	0

On the level "To grip/To loosen", the diagnosis nodes of sub-functions indicate that the failure is due to the sub-function "To actuate". Within this sub-function, the component "JACK" is the cause:

JACK	
OK	0
Jammed	**1**
Blocked	0

Table 2. First inference.

	HD	RWD				
	Raw Part	Gripping	Putting in rotation	Machining	Loosening	Producing
R	0.900	0.854	0.845	0.746	0.655	0.655
NC	0.075	0.056	0.060	0.142	0.227	0.227
NR	0.025	0.088	0.093	0.111	0.117	0.117

5. CONCLUSION AND FURTHER WORK

The proposed model based on the functioning and malfunctioning studies carried out when design makes it possible to build easily the structure of the BN. This structure is based on several levels of abstraction carried out from a generic formalism.

The phase of quantification of the BN is based on the occurrence probability criterion of the FMECA. The translation of the FMECA criterion F makes it possible to obtain prior probabilities. This assessment will have to be updated by the data of operating feedback collected during the exploitation of the system, for example by using Bayesian approach (Suhner, *et al.*, 1997). Thus, the operating feedback will enable to refine the probabilities contained in the CPT by learning.
It is also necessary to associate with all probabilities an uncertainty in order to undertake studies of sensitivity.

The use of the network in order to simulate the impact of a failure on the total performance of the system is easy. The networks can also be used in design to validate the effectiveness of the corrective actions suggested following the FMECA analysis.

During the exploitation of the system, it is effective to use BN for the diagnosis. The strategy described here is only based on the probabilities of occurrence. It is necessary to improve it by taking account other parameters: aptitude of detection, costs of the components…

The taking into account of the economic aspect is also possible in the BN formalism while adding to the network, nodes of costs (e.g. costs related to the detection of a failure) and nodes of decision (e.g. strategy of maintenance).

BNs constitute a powerful tool for decision-making aid in maintenance. The model combines a priori approach in design and a posteriori approach in operation.

It remains to improve the automatic generation of BNs starting from model SADT and FMECA analysis. It is also necessary to validate the model by applying it to a real system in order to show industrial feasibility and to confirm its added value compared to the traditional computerised decision-making systems.

REFERENCES

Bobbio, A., L. Portinale, M. Minichino and E. Ciancamerla (2001). Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliability Engineering and System Safety*, **71**, 249-260.

Bouissou, M., F. Martin and A. Ourghanlian (1999). Assessment of a Safety-Critical System Including Software: A Bayesian Belief Network for Evidence Sources. *RAMS'99 Reliability and Maintainability Symposium*. Washington.

Hu, W., A.G. Starr, Z. Zhou and A.Y.T. Leungt (2000). A systematic approach to integrated fault diagnosis of flexible manufacturing systems. *International Journal of Machine Tools and Manufacture*, **40**, 1587-1602.

Jensen, F.V. (1996). *An Introduction to Bayesian Networks*. UCL Press. London.

Kang, C.W. and M.W. Golay (1999). A Bayesian belief network-based advisory system for operational availability focused diagnosis of complex nuclear power systems. *Expert Systems with Applications*, **17**, 21-32.

Khoo L.P., C.L. Ang and J. Zhang (2000). A fuzzy based genetic approach to the diagnosis of manufacturing systems. *Engineering Applications of Artificial Intelligence*, **13**, 303-310.

Léger, J-B., B. Iung, A. Ferro De Beca and J. Pinoteau (1999). An innovative approach for new Distributed Maintenance System: application to Hydro Power Plants of the REMAFEX project. *Computers in Industry*, **38**, issue 2, 133-150.

Léger J-B. and B. Iung (1998). Methodological approach to modelling of degradation detection and failure diagnosis in complex production system. In: *Proceedings of the ninth international workshop on Principles of Diagnosis DX98*, 209-216, Cap Cod, Massachusetts, May 24-27.

Suhner, M.-C., D. Larcher and M. Gabriel (1997). Utilisation de l'AMDEC pour formaliser le retour d'expérience en utilisant une approche bayésienne. *CANCAM'97, 16ème congrès canadien de mécanique appliquée*, 585-586, Québec (Canada), 1er-5 juin.

Suhner, M.-C., M. Gabriel and L. Claude (1992). Use of FMECA for fault diagnosis of complex automotive systems. *The European Safety and Reliability'92 Symposium*, 598-606, Copenhague (Danemark), June 10-12.

Weber, P. and M.C. Suhner (2001). System architecture design based on a Bayesian Networks method. *Xth International Symposium on Applied Stochastic Models and Data Analysis*, Compiègne, France, June 12-15, 2001.