Bayesian network for the characterization of faults in a multivariate process

Sylvain Verron, Teodor Tiplica, Abdessamad Kobi

Bayesian network for the characterization of faults in a multivariate process

Sylvain VERRON, Teodor TIPLICA, Abdessamad KOBI

LASQUO/ISTIA, Angers - France

sylvain.verron@univ-angers.fr, teodor.tiplica@univ-angers.fr, abdessmad.kobi@univ-angers.fr

Abstract

The main objective of this paper is to present a new method of detection and characterization with a bayesian network. For that, a combination of two original works is made. The first one is the work of Li et al. [1] who proposed a causal decomposition of the T² statistic. The second one is our previous work on the detection of fault with bayesian networks [2], [3], notably on the modelization of multivariate control charts in a bayesian network. Thus, in the context of multivariate processes, we propose an original network structure allowing deciding if a fault is appeared in the process. More, this structure permits the identification of the variables that are responsible (root causes) of the fault. A particular interest of the method is the fact that the detection and the identification can be made with a unique tool: a bayesian network.

1. INTRODUCTION

Nowadays, industrial processes have more and more sensors, giving an important amount of data. An interesting research field is the use of this data in order to control the process. The process control can be view as a four steps procedure [4]. In the first step, the fault detection, the objective is to detect an abnormal situation, a fault in the process. The goal of the second stage, the fault identification, is to identify the most significant variables for the diagnosis of the fault. The third step is the fault diagnosis, this step gives which type of fault is appeared in the process. Finally, the last step is the process recovery which allow to act on the process in order to retrieve the normal operating conditions.

The process control can be made by two main approaches [4]: the analytical approach and the data-driven approach. The analytical techniques are based on analytical (physical) models of the system and enable to simulate the system. Though, at each instant, the theoretical value of each sensor can be known for the normal operating state of the system. As a consequence, it is relatively easy to see if the real process values are similar to the theoretical values. But, the major drawback of this family of technique is the fact that a detailed model of the process is required in order to control it efficiently. An effective detailed model can be very difficult, time consuming and expensive to obtain, particularly for large-scale systems with many variables. The data-driven approaches are a family of different techniques based on the analysis of the real data extracted from the process. These methods are based on rigorous statistical development of the process data (i.e. control charts, methods based on Principal Component Analysis, Projection to Latent Structure or Discriminant Analysis) [4].

In the literature, one can find a lot of data-driven techniques for the process control: univariate statistical process control (Shewhart control chart) [5], multivariate statistical process control (T^2 and Q control charts) [6], and some techniques based on Principal Component Analysis [7] like the multiway PCA or the dynamic PCA [8]. Kano et al. [9] have made a comparison of these different techniques. About the fault identification, Tiplica et al. [10] have given a comparison of several methods. One of the most interesting statistical technique is the MYT decomposition [11] which makes a decomposition of the T^2 statistic in orthogonal components allowing to determine which variable or group of variables has contributed to the out-of-control situation (fault). Recently, Li et al. [1] have proposed an improvement to the MYT decomposition: the causal decomposition of the T^2. In order to make this decomposition, authors use a causal Bayesian network representing the different variables of the process.

As we have said, detection and identification of faults are based on different tools (control chart, various decomposition, …). It will be interesting to obtain all these techniques solely in one tool. The objective of this article is to propose an improvement to the decomposition method of Li et al. [1], in order to use a one and only bayesian network, for the detection of a fault and also in order to identify the implicated variables.

The article is structured as follows: the second section highlights some aspects on bayesian networks and particularly on bayesian network classifiers; the third section presents the various decomposition (causal and MYT); in the fourth section we are reminding how to model some multivariate control charts in a bayesian network, and we present how to obtain detection and identification of faults in a sole Bayesian network; an example of the approach is given on a simple process in the fifth section; in the last section, we conclude on the proposed approach and give some outlooks.

2. BAYESIAN NETWORKS

An interesting tool using statistics is bayesian network, an oriented probabilistic graphic model. Bayesian networks can be efficient supervised classifiers. A bayesian network [12] is a triplet {G, E, D} where:

{G} is a directed acyclic graph, G=(V,A), where V is the set of nodes of G, and A is the set of edges of G,

{E} is a finite probabilistic space (Ω, Z, P), where Ω is a non-empty space, Z is a collection of subspace of Ω, and P is a probability measure on Z with P(Ω)=1,

{D} is a set of random variables associated to the nodes of G and defined on E such as:

$$P(V_1,V_2,\ldots,V_n)=\prod_{i=1}^{n} P(V_i | C(V_i)) \quad (1)$$

where C(Vi) is the set of parents of Vi in the graph G.

Bayesian network classifiers are particular bayesian networks [13]. They always have a discrete node C coding the k different classes of the system. The remaining variables X_i represent the descriptors (variables) of the system.

A Naïve Bayesian Network (NBN) is a particular type of bayesian network classifiers. It is also known as the Bayes classifier. In a NBN, the class node is linked with all other variables of the system (descriptors) as indicated on the figure 1.

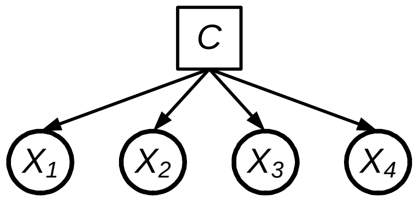

Fig. 1 Example of a Naïve Bayesian Network (NBN).

The NBN is called naïve because it makes the naïve (but strong) assumption that all descriptors (variables of the system) are class conditional statistically independent (no correlation between each descriptor in each class). But, in many systems, it is very frequent to have high correlations between variables, and a NBN will not take into account these correlations. Extensions of NBN have been developed in order to solve this problem.

A first interesting extension is the TAN (Tree-Augmented bayesian Network) [13]. In a TAN, a maximum weighted spanning tree is constructed with the descriptors following the algorithm of Chow and Liu [14]. So, each descriptor will have at most one other descriptor as parent. After that, edges from the class node to each descriptor are added (like a NBN). An example of a TAN is given on the figure 2.

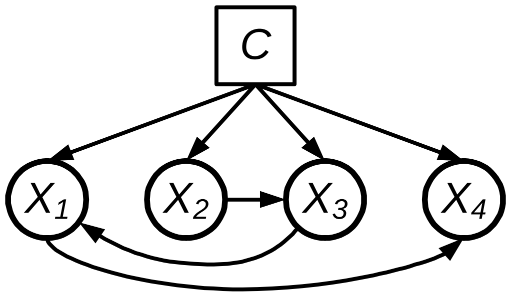

Fig. 2 Example of a Tree-Augmented bayesian Network (TAN).

Finally, the bayesian multinets [13] introduce a different structure for each value of the class variable (a particular case is to take a different TAN for each value of the class). In [13], authors show that these classifiers give a lower misclassification rate than the NBN. But, these classifiers do not take into account all the correlations between variables.

An other extension to the NBN is the Condensed Semi Naïve Bayesian Network (CSNBN) [15]. The principle of this classifier is to represent some variables in a joint node (i.e. some normally distributed variables can be modeled with a node representing a multivariate normal distribution). In this way, all correlations of the system will be taken into account. A CSNBN will be composed of two nodes: the class node and a multivariate node. An example of a CSNBN is given on the figure 3.

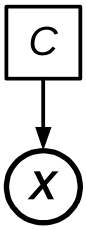

Fig. 3 Example of a Condensed Semi Naïve Bayesian Network (CSNBN).

3. Different Methods

3.1. MYT decomposition

As we previously said, a method for the fault detection in multivariate processes is the T^2 control chart. However, this chart do not give any information about the diagnosis of the out-of-control situation. For that, many techniques have been proposed in the literature [11], [16], [17]. A comparative study has been made by Tiplica [18].

An interesting decomposition of the T^2 has been proposed by Mason, Young et Tracy [11], namely “MYT decomposition. More, in order to better understand this method, authors give an example for a bivariate process. One can be precise that the authors have proved that several methods can be considered like special cases of MYT decomposition [11].

The principle of this method is to decompose the T^2 statistic in a limited number of orthogonal components which are also statistical distance (and so, can be monitored). This decomposition is the following:

$$T^2 = T_1^2 + T_{2\bullet 1}^2 + T_{3\bullet 1,2}^2 + T_{4\bullet 1,2,3}^2 + \cdots + T_{p\bullet 1,2,3\cdots p-1}^2 \quad (2)$$

where $T_{i\bullet j,k}^2$ represents the T^2 statistic of the regression of the variables X_j and X_k on the variable X_i. We remark that it exists a large number of different decompositions (p!) and so, it exists a large number of different terms ($p(2^{p-1})$). In order to better understand, on a process with three variables, the different decompositions available are:

$$\begin{aligned} T^2 &= T_1^2 + T_{2\bullet 1}^2 + T_{3\bullet 1,2}^2 \\ T^2 &= T_1^2 + T_{3\bullet 1}^2 + T_{2\bullet 1,3}^2 \\ T^2 &= T_2^2 + T_{1\bullet 2}^2 + T_{3\bullet 1,2}^2 \\ T^2 &= T_2^2 + T_{3\bullet 2}^2 + T_{1\bullet 2,3}^2 \\ T^2 &= T_3^2 + T_{1\bullet 3}^2 + T_{2\bullet 1,3}^2 \\ T^2 &= T_3^2 + T_{2\bullet 3}^2 + T_{1\bullet 2,3}^2 \end{aligned} \quad (3)$$

The computation of the different terms is not detailed in this article, but we report the readers to the works of Mason et al. [11]. We can note that the terms T_j^2 are called non conditional terms and that the other terms are called conditional terms. Each terms follows a Fisher distribution law:

$$T_{j+1\bullet 1,\cdots j}^2 = \frac{(m+1)(m-1)}{m(m-k-1)} F_{1,m-k-1} \quad (4)$$

where k is the number of conditioned factors. We can simplify this equation for the non conditional terms (k=0) by:

$$T_{j+1\bullet 1,\cdots j}^2 \sim \frac{m+1}{m} F_{1,m-1} \quad (5)$$

This monitoring allows to detect a problem on each term of the decomposition. For example, if we see that the term $T_{2\bullet 1}^2$ is responsible of the out-of-control of the process, we can immediately search a root cause on a tuning of the process touching to the correlation between these two variables. But for a more simplest computation, one can use a T^2 control chart for the detection of the fault, and in the case of a detected fault we can apply the MYT decomposition.

The main advantage of the MYT decomposition is the fact that this method can give a diagnosis of the situation without any samples of previous faults. An other advantage is the fact that this method is based on the same statistics tools that the T^2 control chart.

3.2. Causation-based T^2 decomposition

The MYT method is very interesting, but it has a major drawback: the number of term to compute. Indeed, as we previously said, the number of terms to compute is equal to $p(2^{p-1})$. For example, for a process with 20 variables, more than 10 millions of terms are needed to compute. A five steps algorithm has been proposed by Mason et al. [11] in order to reduce the number of terms to compute. But, Li et al. [1] said that the number of terms is always too large. So, Li et al. [1] propose a new method with a bayesian network: the causation-based T^2 decomposition. A causal graph which represents the process allows to reduce the number of terms to only p terms to compute. More than the decrease of computation giving by this method, the authors show that one obtain equally an increasing of the performances.

The basis assumption of the method proposed by Li et al. [1] is that the process can modelized with a causal bayesian network where each variable of the process is a gaussian univariate variable. If a bayesian network represents solely some gaussian continuous variables, it is also called linear gaussian model. So, for a process with 3 variables, we can obtain, for example, the network of the figure 4.

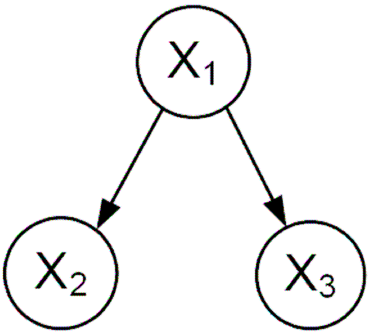

Fig. 4 Example of a linear gaussian model.

Concerning the modelization of the process by a linear gaussian model, the authors make the distinction between two types of decomposition MYT: “for a given decomposition of the T², if it exists a term $T^2_{i\bullet 1,\ldots,i-1}$ such the ensemble of the variable $\{X_1, \ldots, X_{i-1}\}$ have at least one descendant of X_i, then this decomposition is a type A decomposition, if not, the decomposition is a type B decomposition. So, we can sort in the table 1 the different decomposition of the process with 3 variables of the figure 4.

Décomposition	Type
$T^2 = T_1^2 + T^2_{2\bullet 1} + T^2_{3\bullet 1,2}$	Type B
$T^2 = T_1^2 + T^2_{3\bullet 1} + T^2_{2\bullet 1,3}$	Type B
$T^2 = T_2^2 + T^2_{1\bullet 2} + T^2_{3\bullet 1,2}$	Type A
$T^2 = T_2^2 + T^2_{3\bullet 2} + T^2_{1\bullet 2,3}$	Type A
$T^2 = T_3^2 + T^2_{1\bullet 3} + T^2_{2\bullet 1,3}$	Type A
$T^2 = T_3^2 + T^2_{2\bullet 3} + T^2_{1\bullet 2,3}$	Type A

Tab. 1. Decomposition types for the three variables process.

Li et al. [1] proved, basing on the Hawkins works [16], that the type A decompositions allow a less accurate diagnosis than the type B decompositions. More, they proved that in the context of linear gaussian models, all the type B decomposition converge to a sole decomposition that the authors named “causation-based T² decomposition”. Indeed, each type B decomposition converge to the causal decomposition of the T² given in equation 5, where PA_{Xi} represents the parents of the variable X_i in the causal graph.

$$T^2 = \sum_{i=1}^{p} T^2_{i\bullet PA(X_i)} \qquad (6)$$

So, the causation-based T² decomposition of the example of the figure 4 is the following :

$$T^2 = T_1^2 + T^2_{2\bullet 1} + T^2_{3\bullet 1} \qquad (7)$$

After that, the authors give the procedure of detection and identification of fault using the new causal decomposition. Firstly, a linear gaussian bayesian network is constructed in order to represent the different causal relations between the process variables. Secondly, the process is monitored with a T² control chart. In the case of an out-of-control situation, the T² is decomposed by the causation-based T² decomposition of the equation (6). In this equation, each $T^2_{i\bullet PA(X_i)}$ is independent and, in the case of known parameters, follows a χ^2 distribution law with one degree of freedom. So, each $T^2_{i\bullet PA(X_i)}$ can be compared to the limit $\chi^2_{1,\alpha}$ which is the quantile at the value α (α is the false alarm rate) of the χ^2 distribution with one degree of freedom. A significant term $T^2_{i\bullet PA(X_i)}$ (higher than the limit) give that the variable X_i has been implicated in the fault. The figure 5 represents the process monitoring diagram with the given method.

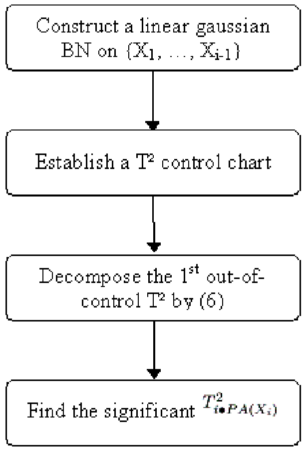

Fig. 5 The process monitoring diagram.

The approach developed by Li et al. [1] exploits some threshold given by quantiles of statistical laws. This method allows to considerably increase the performances compared to the MYT method, with less computation resource. But, we can view on figure 5 that several tools are used for this technique: control chart, bayesian network, statistical computations. We will show, in the next section, that the monitoring of a multivariate process by the method of the causal-based T² decomposition can entirely realised in a sole bayesian network.

4. Proposed Approach

4.1. Control charts with bayesian network

In previous works [2], [3], we have demonstrated that a T² control chart [19] could modelized with a bayesian network. For that, we use two nodes: a gaussian multivariate node **X** representing the data and a bimodal node E representing the state of the process. The bimodal node E has the following modalities: IC for in control and OC for out-of-control. Assuming that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are respectively the mean vector and the variance-covariance matrix of the process, we can monitor the process with the following rule: if P(IC|x)<P(IC) then the process is out-of-control. This bayesian network is represented on the figure 6, where the conditional probabilities tables for each node are given.

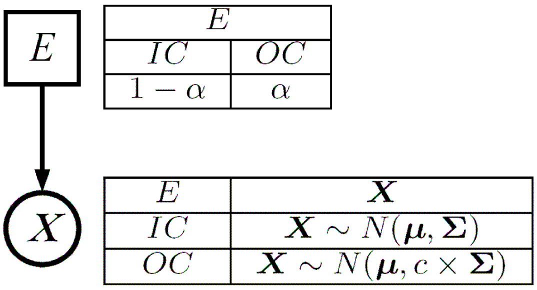

Fig. 6 T^2 control chart in a bayesian network.

On this figure 6, we can observe that a coefficient c is implicated in the modelization of the control chart by bayesian network. This coefficient is the root (different of 1) of the following equation:

$$1-c+\frac{pc}{CL}\ln(c)=0 \quad (8)$$

where p is the dimension (number of variables) of the system to monitor, and CL is the control limit of the equivalent T^2 control chart. In numerous cases, CL is equal to $\chi^2_{p,\alpha}$, the quantile at the value α of the distribution of the χ^2 with p degree of freedom [20]. So, α allows to tune the false alarm rate of the control chart.

4.2. Improvement of the causal decomposition

The method proposed by Li et al. [1] allows, based on a Bayesian network, to know the different terms of the MYT decomposition to compute. For the computation of the different terms of the causation-based T^2 decomposition, and for the associated decisions (use of the threshold), the authors do not use the network in an optimal way. Indeed, they use a T^2 control chart out of the network, but this chart can be modelized directly in the network (see previous section). More, the authors compute each $T^2_{i\bullet PA(X_i)}$ out of the network, but it is possible to make all the computations in the network.

We propose an extension to the method of Li et al. [1] allowing the computation of the different $T^2_{i\bullet PA(X_i)}$ and the decisions associated to each one. The diagnosis by the causation-based T^2 decomposition, like the MYT decomposition, is a monitoring of regressed variables, with the use of univariate control charts. In the previous section, we have demonstrated how to realize, in a Bayesian network, a multivariate control chart like the T^2 control chart. But, a univariate control chart like a Shewhart control chart is simply a particular case of a multivariate control chart like the T^2 control chart. Indeed, the computation of the T^2 is the following :

$$T^2 = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \quad (9)$$

But, in the univariate case, the previous equation gives :

$$T^2 = \frac{(x-\mu)^2}{\sigma^2} \quad (10)$$

In this univariate case, the T^2 statistic follows a χ^2 distribution law with one degree of freedom. So, it is possible to improve the method developed by Li et al. [1]. We propose to monitor directly the different values of the $T^2_{i\bullet PA(X_i)}$ in the bayesian network. For that, we add a bimodal variable to each univariate node of the Bayesian network. If we have a graph representing a system with three variables (see figure 4), we then obtain a network with six nodes: 3 continuous (gaussian univariate) and 3 discrete (bimodal), like shown on figure 7.

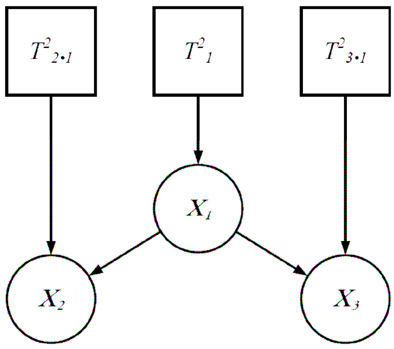

Fig. 7 Improvement of the causal decomposition.

We precise here that the continuous variables are not needed to be standardized before. The discrete nodes added to the initial structure of the network allow to directly do the identification of responsible variables in an out-of-control situation. These nodes modelize a control chart $T^2_{i\bullet PA(X_i)}$ allowing to conclude on the state of each variable of the process. We recall that the modality IC of a discrete node signify in control, and that the modality OC signify out-of-control. The figure 8 gives the different probabilities tables associated to the different node (discrete or continuous).

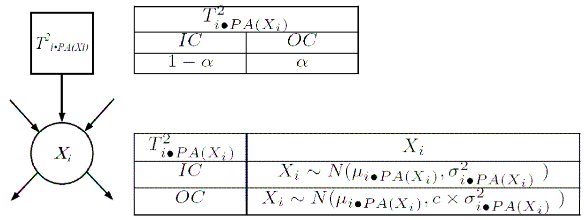

Fig. 8 Bayesian network part.

When a fault is detected in the process, each discrete node (representing the status of a regressed variable) give a certain probability that the variable is in control. The responsible variables are ones with an a posteriori probability lower than the a priori probability. The operator can then easily search for the root cause (the physical cause) because he knows which variables are responsible of the out-of-control status.

To the view of all our statements, we can propose a Bayesian network able to determine if the process is in or out-of-control (detection). In the case of a out-of-control situation, this bayesian network can also gives the implicated variables (identification). The figure 9 presents the general form of the bayesian network for a process with p variables. This sole network is able to do all the step of the diagram of the Li et al. methods.

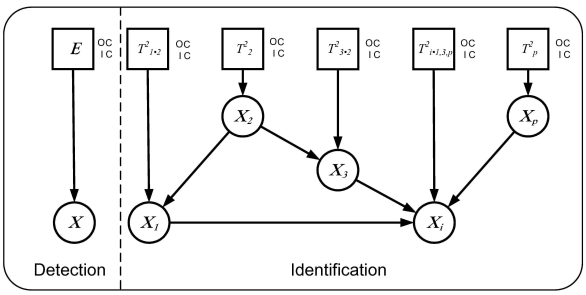

Fig. 9 Bayesian network for the detection and the identification.

5. APPLICATION

In order to better understand the proposed method, we have simulated a process with three variables like presented on figures 4 and 6. Parameters **μ** and **Σ** of this process, in the in control case, are the following:

$$\boldsymbol{\mu} = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$$
$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.8 & 0.6 \\ 0.8 & 1 & 0.1 \\ 0.6 & 0.1 & 1 \end{pmatrix} \quad (11)$$

We have simulated 100 in control observations, and 50 observation of out-of-control. These out-of-control observations are a mean shift of magnitude 2 on the variable X_2. The objective of the network is to detect that a fault is appeared from the observation 101, and that the variable X_2 is implicated in this fault.

The first step is the construction of the network. For that, we use the PC algorithm [21]. For this algorithm, we have used a conditional independencies tests with Fisher's z-transform [22]. Then, we can add all the node for the control of the process, namely all the nodes $T^2_{i \bullet PA(X_i)}$, with a false alarm rate of 5%. We add also the modelization of the T^2 control chart, allowing the detection in the network, with a false alarm rate of 1%. After that, each parameters of the nodes are computed.

As the structure and the parameters of the network have been learnt, we have presented the 150 observations in order to obtain a result for each one. So, we observe the a posteriori probabilities of the node E. For a given observation, if this probability is lower than 99% then the process is out-of-control. In this case, we observe the values of the a posteriori probabilities of the nodes $T^2_{i \bullet PA(X_i)}$. Then, variables with a probability lower than 95% are implicated in the fault. The results (a posteriori probabilities) of the network are given on the figure 10.

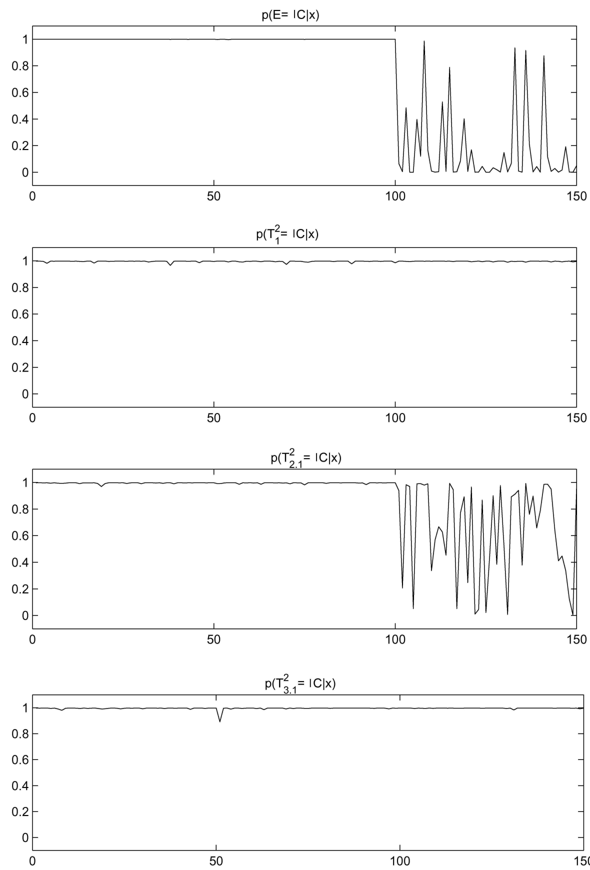

Fig. 10 A posteriori probabilities of discrete nodes.

On this figure, we can see that the network do not detect fault in the 100 first observations. However, the network detects an abnormal situation from the observation 101. More the network said that the solely the variable X_2 is implicated in the fault.

6. CONCLUSIONS AND OUTLOOKS

In this article, we have presented an approach allowing the fault detection and fault identification of a multivariate process. This approach is based on a Bayesian network. We have combined previous works of control chart in Bayesian networks with some recent works of Li et al. [1] . The interesting outlook of this work is the set up of a sole Bayesian network allowing the complete control of the process, including detection, identification and diagnosis.

REFERENCES

[1] **Jing Li, Jionghua Jin** and **Jianjun Shi**: ”Causation-based T² decomposition for multivariate process monitoring and diagnosis”, *Journal of Quality Technology*, 40(1), 2008, pp. 46-58.

[2] **Sylvain Verron, Teodor Tiplica** and **Abdessamad Kobi**: ”Multivariate control charts with a bayesian network”, *4th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*}, 2007, pp 228-233.

[3] **Sylvain Verron**: ”Diagnostic et surveillance des processus complexes par réseaux bayésiens”, PhD thesis, University of Angers, 2007.

[4] **Leo H. Chiang, Evan L. Russell,** and **Richard D. Braatz**,: ”Fault detection and diagnosis in industrial systems”, New York: Springer-Verlag, 2001.

[5] **Walter A. Shewhart**: ”Economic control of quality of manufactured product, New York : D. Van Nostrand Co., 1931.

[6] **J.A. Westerhuis, S.P. Gurden,** and **A.K. Smilde**: ”Standardized q-statistic for improved sensitivity in the monitoring of residuals in MSPC”, *Journal of Chemometrics*, 14(4), 2000, pp. 335 – 349.

[7] **Edward J. Jackson**: ”Multivariate quality control”, *Communication Statistics - Theory and Methods*, 14, 1985, pp. 2657 – 2688.

[8] **Bhavik R. Bakshi**: ”Multiscale PCA with application to multivariate statistical process monitoring”, *AIChE Journal*, 44, 1998, pp. 1596 - 1610.

[9] **M. Kano, K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss,** and **B.R. Bakshi**: ”Comparison of multivariate statistical process monitoring methods with applications to the eastman challenge problem”, *Computers and Chemical Engineering*, 26(2), 2002, pp. 161 – 174.

[10] **Teodor Tiplica, Abdessamad Kobi,** and **Alain Barreau**: “Synthèse et comparaison des méthodes pour la maîtrise statistique des processus multivariés”, *Actes du congès QUALITA*, 2001, pp. 134 – 142.

[11] **Robert L. Mason, Nola D. Tracy,** and **John C. Young**: ”Decomposition of T² for multivariate control chart interpretation”, *Journal of Quality Technology*, 27(2), 1995, pp. 99 - 108.

[12] **Patrick Naim, Pierre-Henri Wuillemin, Philippe Leray, Olivier Pourret,** and **Anna Becker**: ”Réseaux bayésiens - 2ème édition”, Eyrolles, 2004.

[13] **N. Friedman, D. Geiger,** and **M. Goldszmidt**: ”Bayesian network classifiers”, *Machine Learning*, 29(2-3), 1997, pp. 131 - 163.

[14] **C. Chow** and **C. Liu**: ”Approximating discrete probability distributions with dependence trees*”, IEEE Transactions on Information Theory*, 14(3), 1968, pp. 462 – 467.

[15] **Igor Kononenko**: ”Semi-naive bayesian classifier”, *EWSL-91: Proceedings of the European working session on Machine learning*, 1991, pp. 206 – 219.

[16] **Douglas M. Hawkins**: ”Regression adjustment for variables in multivariate quality control”, *Journal of Quality Technology*, 25(3), 1993, pp. 170 – 182.

[17] **Teodor Tiplica, Abdessamad Kobi,** and **Alain Barreau**: ”Optimisation et maîtrise des processus multivariés: la méthode FNAD”, *Journal Européen des Systèmes Automatisés*, 37(4), 2003, pp. 477 – 500.

[18] **Teodor Tiplica**: ”Contribution à la Maîtrise Statistique des Processus Industriels Multivariés”, PhD thesis, ISTIA, 2002.

[19] **Harold Hotelling**: ”Multivariate quality control”, *Techniques of Statistical Analysis*, 1947, pp. 111 - 184.

[20] **Douglas C. Montgomery**: ”Introduction to Statistical Quality Control, Third Edition”, John Wiley and Sons, 1997.

[21] **Peter Spirtes, Clark Glymour,** and **Richard Scheines**: Causation, prediction, and search”, Springer-Verlag, 1993.

[22] **M. Kalisch** and **P. Buhlmann**: ”Estimating high-dimensional directed acyclic graphs with the PC-algorithm”, *Journal of Machine Learning Research*, 8, 2007, pp. 613 - 636.