

Fault detection of univariate non-Gaussian data with Bayesian network

Sylvain Verron, Teodor Tiplica, Abdessamad Kobi

► **To cite this version:**

Sylvain Verron, Teodor Tiplica, Abdessamad Kobi. Fault detection of univariate non-Gaussian data with Bayesian network. IEEE International Conference on Industrial Technology (ICIT'10), 2010, Vina del Mar, Chile. 2010. <inria-00517031>

HAL Id: inria-00517031

<https://hal.inria.fr/inria-00517031>

Submitted on 13 Sep 2010

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.

Fault detection of univariate non-Gaussian data with Bayesian network

Sylvain Verron, Teodor Tiplica and Abdessamad Kobi

LASQUO/ISTIA

62, av. Notre Dame du Lac

49000 Angers, France

Email: sylvain.verron@univ-angers.fr

Abstract—The purpose of this article is to present a new method for fault detection with Bayesian network. The interest of this method is to propose a new structure of Bayesian network allowing to detect a fault in the case of a non-Gaussian signal. For that, a structure based on Gaussian mixture model is proposed. This particular structure allows to take into account the non-normality of the data. The effectiveness of the method is illustrated on a simple process corrupted by different faults.

I. INTRODUCTION

Nowadays, an important topic of research is the Fault Detection and Diagnosis (FDD) ([1]). Classically, FDD is an essential task, it allows: to insure safety conditions for humans and materials, to improve quality products, to reduce manufacturing cost, and so on. Generally, FDD is composed of two different steps: the fault detection and the fault diagnosis. The objective of the fault detection step is to decide if a fault has appeared in the process. Once a fault has been detected, the goal of the fault diagnosis step is to find the root cause of the detected fault.

In the literature, we can find two main approaches for the FDD. The first one is the model-based approach ([2]). This first approach consist to represent the process in an analytical form ([3]). For that, physical equations describing the process are taken into account in order to model the process. Once an efficient model of the process is found, some methods can be used in order to generate residuals ([4]). A residual corresponds to the difference between the measurement of a variable in the process and his theoretical value given by the analytical model. So, if a residual is close to 0, we can consider that no fault affect this variable. Otherwise (residual not equal to 0), we can consider that a fault has appeared in the process (fault detection). Based on the different symptoms (significant residuals), a diagnosis of the appeared fault can be made with the help of an incidence matrix ([5]).

The second approach for the FDD is the data-driven approach ([6]). This approach makes the assumption that some data are available from the process. Generally, fault detection with data-driven methods requires only some data from normal operating conditions (fault-free). For the fault diagnosis with data-driven, we can distinguish two approaches, which can be complementary. The first one is the characterization of the fault. This approach consists to conclude on which variables are implicated in the fault ([7]). The second approach is a

supervised classification task. Based on the data given in previous faults, the objective is to allocate a class (type of previous fault) to a new detected fault ([8]). In this case, many classifiers can be used: FDA (Fisher Discriminant Analysis) ([9]), SVM (Support Vector Machine) ([10]), kNN (k-nearest neighborhood) ([11]), ANN (Artificial Neural Networks) ([9]) and Bayesian networks classifiers ([12]).

In recent years, some works have been proposed in order to achieve the FDD with the help of Bayesian networks ([13], [14], [15], [16]). Bayesian networks have particular advantages for the Fault Detection and Diagnosis. Indeed, this tool can handle discrete and continuous variables, and can describe probabilistic relations between these variables. Moreover, some extensions of Bayesian networks seem to be really useful for the FDD, like dynamic Bayesian networks ([17]), object-oriented Bayesian networks ([18]) or influence diagrams ([9]).

A Bayesian Network (BN) ([19]) is an acyclic graph where each variable is a node (that can be continuous or discrete). Edges of the graph represent dependences between linked nodes. A formal definition is given here:

A Bayesian network is a triplet $\{\mathbf{G}, \mathbf{E}, \mathbf{D}\}$ where:

- $\{\mathbf{G}\}$ is a directed acyclic graph, $\mathbf{G} = (V, A)$, with V the ensemble of nodes of \mathbf{G} , and A the ensemble of edges of \mathbf{G} ,
- $\{\mathbf{E}\}$ is a finite probabilistic space (Ω, Z, p) , with Ω a non-empty space, Z a collection of subspace of Ω , and p a probability measure on Z with $p(\Omega) = 1$,
- $\{\mathbf{D}\}$ is a set of random variables associated to the nodes of \mathbf{G} and defined on \mathbf{E} such as:

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

with $C(V_i)$ the ensemble of causes (parents) of V_i in the graph \mathbf{G} .

Some methods have already been proposed for the case of a Gaussian signal ([20]). But, signals from industrial processes are sometimes non-Gaussian. For example, in mechanics, we can found the concentricity error or the symetry error. So, the purpose of this article is to present a new method for the detection of faults in industrial systems in the case where the fault-free data are non-Gaussian.

The article is structured in the following manner. In section II, we present the method allowing the fault detection of Gaussian signal with Bayesian network. The section IV presents a new method in order to take into account non-Gaussian signal in the fault detection with Bayesian network. We evaluate this proposed method on a simple example in section V. Finally, section VI concludes on interests and limitations of this method, and presents some perspectives of the Fault Detection and Diagnosis with Bayesian networks.

II. BN FOR FAULT DETECTION IN THE GAUSSIAN CASE

Bayesian networks are able to directly handle Gaussian variables (univariate and multivariate). In the FDD framework, one wants to detect if a signal is stable or not. In the Gaussian case, we can view the problem as in the figure 1.

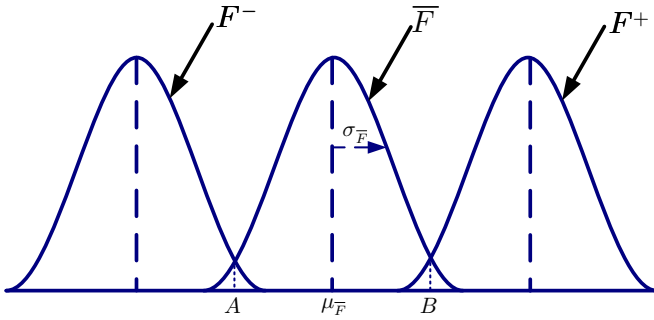


Fig. 1. Illustration of the fault detection statement

If the signal is stable (no fault), the variable X follows a Gaussian distribution \bar{F} of parameters $\mu_{\bar{F}}$ and $\sigma_{\bar{F}}$. If a positive shift affects the signal, X will follow a new and supposed unknown distribution F^+ . In the same way, if a negative shift affects the signal, X will follow a new and supposed unknown distribution F^- . Classically, the following rule is used: if x is comprised between A and B , then we accept the assumption that X follows \bar{F} . A and B are threshold allowing to respect a given false alarm rate α (or $\alpha/2$ on each side). Generally, A and B are computed with the following equations, where Z represents the quantile of the standard normal distribution:

$$A = \mu_{\bar{F}} + Z_{\frac{\alpha}{2}} \sigma_{\bar{F}} \quad (2)$$

$$B = \mu_{\bar{F}} - Z_{\frac{\alpha}{2}} \sigma_{\bar{F}} \quad (3)$$

Detection for the Gaussian case will be to decide if X follows \bar{F} , F^- or F^+ . In the case of F^- or F^+ , we have a detection (a change in the signal of X has been detected). So, the detection can be seen as a discriminant analysis. Indeed, for a given observation x , the fault detection is similar to the following question: is x allocated to \bar{F} , F^- or F^+ ? In a Bayesian network, the application of a discriminant analysis can be done with a simple structure (in the univariate case), as given on the figure 2: a discrete variable C models the classes of the problem, thus a continuous (Gaussian) variable X models the signal.



Fig. 2. Structure of the discriminant analysis

In a discriminant analysis, the classical rule is to allocate the observation x to the class with the highest a posteriori density. So, on the figure 1, we can see that we have to execute a discriminant analysis with 3 classes, and that we have to fix the different parameters of the networks in order to respect the fact that the density of \bar{F} must be greater than the two others (densities of F^- and F^+) between A and B .

As the distribution of F^- and F^+ are supposed unknown, we model them by Gaussian distributions. But, we will impose parameters for these distributions allowing the already enounced rule: if x is comprised between A and B , then we accept the assumption that X follows \bar{F} . For that, we fix σ_{F^-} and σ_{F^+} (respectively the standard deviations of F^- and F^+) to $\sigma_{\bar{F}}$. So, if $\sigma_{\bar{F}} = \sigma_{F^-} = \sigma_{F^+}$, then in order to respect the detection rule, we have to fix μ_{F^-} and μ_{F^+} (respectively the means of F^- and F^+) as follow:

$$\mu_{F^-} = \mu_{\bar{F}} + 2Z_{\frac{\alpha}{2}} \sigma_{\bar{F}} \quad (4)$$

$$\mu_{F^+} = \mu_{\bar{F}} - 2Z_{\frac{\alpha}{2}} \sigma_{\bar{F}} \quad (5)$$

Finally, we obtain the following Bayesian network and conditional probabilities tables (figure 3).

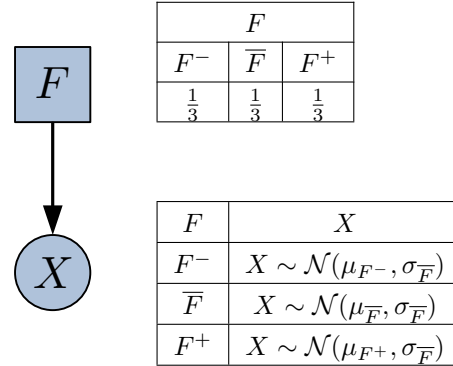


Fig. 3. BN for fault detection in Gaussian signal

The input of this network is the data from the variable X . Once this data is entered as observation (evidence), inference gives the a posteriori probabilities of each class (\bar{F} , F^- and F^+). For a given observation, the class with an a posteriori probability higher than 0.5 will be chosen. So, 0.5 represents a threshold for the decision about the fault detection: if $P(\bar{F}|x) < 0.5$ then a fault has appeared in the process, else no fault is detected. The previous structure and parameters model

a fault detection scheme in the case of a Gaussian signal. But, in certain cases, the signal to monitor is not Gaussian. So, in these cases, the previous method is not applicable. In the next section, we propose to develop a new method allowing to take into account non Gaussian signal for the fault detection.

III. THE NON-GAUSSIAN CASE

A. Presentation of the non-Gaussian case

As mentioned in the introduction, sometimes in industrial processes, the distribution of the normal (no fault) conditions data is not Gaussian. For example, the probability density function of the signal of the fault-free case can be this of figure 4.

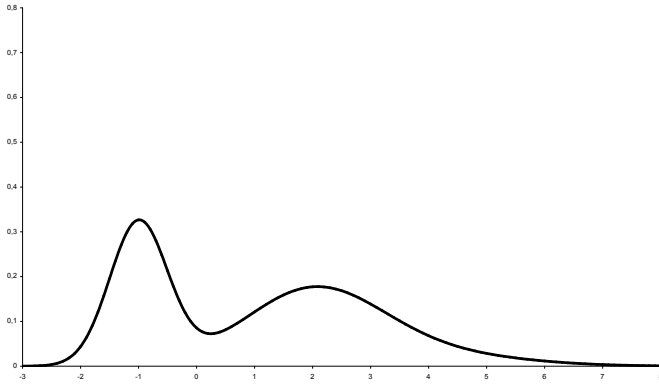


Fig. 4. Example: probability density function of data in the non-Gaussian case

The main problem of Bayesian network is that this tool is not able to directly handle other distribution than the Gaussian one. So, the distribution described in figure 4 cannot be directly model in the network. Some solutions have been proposed to handle various distributions in a Bayesian network: mixture of truncated exponentials ([21]) or kernels methods ([22]). But these types of methods implicate some difficulties for the exact inference, and software are not really exploitable. For these reasons, we prefer to use an other method, directly applicable in classic Bayesian networks (with discrete and Gaussian nodes): the Gaussian Mixture Model (GMM).

B. Gaussian mixture model

A well known semi-parametric method for the estimation of non-Gaussian signal is the Gaussian Mixture Model ([23]). With this method, the probability density function $p(X)$ is seen as a weighted addition of several Gaussian probability density functions. $p(X)$ can be defined as:

$$p(x) = \sum_{j=1}^d a_j p(x|D_j) \quad (6)$$

In this equation, d represents the number of components of the model, and D_j represent the label of the j^{th} component. So, d is the number of Gaussian distributions used for the approximation of the probability density function $p(X)$. a_j

represents the weight allocated to the Gaussian law represented by $p(X|D_j)$. Evidently, for all j , we must have $0 < a_j < 1$. Moreover, one must verify that $\sum_{j=1}^d a_j = 1$.

On the figure 5, we can see that $p(X)$ is the weighted sum of three Gaussian probability density functions: $p(X|D_1)$, $p(X|D_2)$ and $p(X|D_3)$.

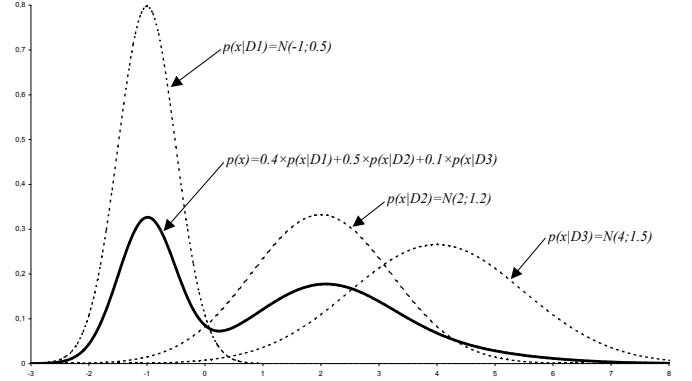


Fig. 5. Example: GMM of the signal of figure 4

Of course, the main difficulty is to obtain the different parameters of the functions $p(X|D_j)$. For that, several methods exist: the Expectation-Maximization (EM) algorithm ([24]), the Markov chains Monte Carlo ([25]), or spectral methods ([26]). The inconvenience of these methods is that they take into account that the number of components d of the model is known. But, in many cases, this number is not known. The main problem is then to compute the optimal value of d . Unfortunately, at present time, a universal solution does not exist for this problem. So, a practical method is to proceed by iteration: take $d = 1$, $d = 2$, and so on, and then do a statistical test between each iteration to see if the increasing of d give or not an increase to the model accuracy.

C. Gaussian mixture model in a Bayesian network

In a Bayesian network, modeling a Gaussian mixture model will be simple. Indeed, if all the parameters of the GMM is known, this can be modeled with two nodes: a discrete one (D) and a Gaussian one (X), as described on the figure 6.

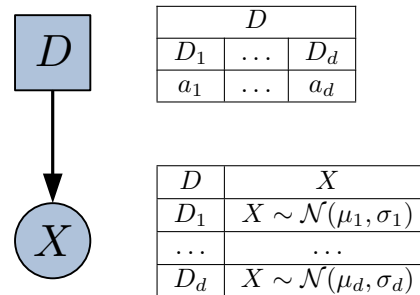


Fig. 6. Example: GMM in a Bayesian network

On the figure 6, the dimension of D represents the number of components d of the GMM. The probability table of D

represents the different coefficients (weights) a_j of the GMM. The conditional nodes X give the d different distribution laws representing the GMM.

IV. BN FOR FAULT DETECTION IN THE NON-GAUSSIAN CASE

In the section II, we have seen how to detect a fault with a Bayesian network for the case of a Gaussian signal. We presented, in section III, how to model a GMM in a Bayesian network. Now, we will propose a method allowing to detect a fault in a Bayesian network when the signal to monitor is not Gaussian.

We first establish the problem statement of the fault detection in the non-Gaussian case. This problem is quite similar to those of section II, but here the signal of the fault-free condition is not Gaussian. So, we have the statement represented in figure 7.

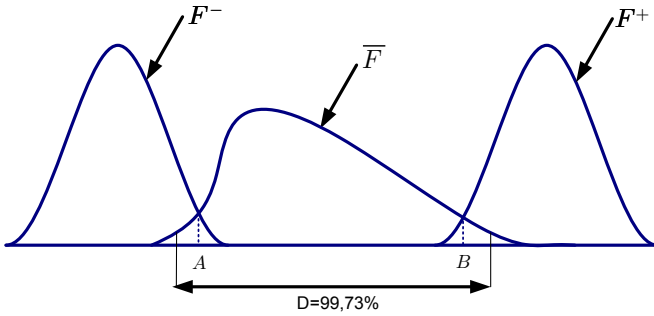


Fig. 7. Illustration of the fault detection statement in the non-Gaussian case

As in the Gaussian case, if the signal is stable (no fault), the variable X follows a distribution called \bar{F} . If a positive shift affects the signal, X follows a new and supposed unknown distribution F_+ , and similarly F_- in the case of a negative shift. A and B are threshold allowing to respect a given false alarm rate α . But here, in the case of a non-Gaussian distribution, the false alarm rate is not necessary divided in two equal parts for each side, we can choose an other repartition of α . Moreover, we introduce D , a dispersion measure, described as representing 99.73% of the population (0.135% on each side).

As in the Gaussian case, the fault detection will be assimilated to a discriminant analysis between the three classes (\bar{F} , F^- and F^+). Moreover, as the fault classes (F^- and F^+) are unknown, we will simply use Gaussian distributions for these two classes (as in the Gaussian case). The more important is to define these two classes in order to keep the decision rule: if x is comprised between A and B , then we accept the assumption that x is allocated to \bar{F} . The different steps giving the definition of the Bayesian network able to detect a fault in a non-Gaussian case are given here.

Firstly, based on the process data in a fault-free case, we estimate the different parameters of the GMM defining the class \bar{F} . For that, we use the EM algorithm. So, we

obtain the parameters of the d Gaussian components of \bar{F} : $\mu_{\bar{F}1}, \sigma_{\bar{F}1}, \dots, \mu_{\bar{F}d}, \sigma_{\bar{F}d}$.

Secondly, as we know the distribution of \bar{F} , we can compute A and B . For that, we use the following conditions: $\Phi_{\bar{F}}(A) = \frac{\alpha}{n}$ and $\Phi_{\bar{F}}(B) = 1 - \frac{\alpha}{n}$, where $\Phi_{\bar{F}}(x)$ represents the cumulative distribution function of \bar{F} , and n is the repartition coefficient of the false alarm rate α .

Thirdly, we compute the parameters of the fault classes (F^- and F^+). We have defined these two classes to follow Gaussian distributions. In order to obtain reasonable variance for these classes compared to the fault-free case, we choose empirically the following standard deviation: $\sigma_{F^-} = \sigma_{F^+} = \frac{D}{6}$. As the standard deviation is fixed, we can compute μ_{F^-} and μ_{F^+} in order to respect $\Phi_{\bar{F}}(A) = \Phi_{F^+}(A)$ and $\Phi_{\bar{F}}(B) = \Phi_{F^+}(B)$. For that, μ_{F^-} is the lower root of the following equation:

$$\mu_{F^-}^2 - 2A\mu_{F^-} + 2A\sigma_{F^-}^2 \log(\Phi_{\bar{F}}(A)\sigma_{F^-}\sqrt{2\pi}) = 0 \quad (7)$$

And, equivalently, μ_{F^+} is the higher root of the following equation:

$$\mu_{F^+}^2 - 2B\mu_{F^+} + 2B\sigma_{F^+}^2 \log(\Phi_{\bar{F}}(B)\sigma_{F^+}\sqrt{2\pi}) = 0 \quad (8)$$

Finally, we can construct the Bayesian network. This network is composed of three nodes. The first one is F and represent the status of the signal, or equivalently the classes of the system. This node has three modalities: \bar{F} , F^- and F^+ . This node is a root node, it has no parent nodes. His a priori probability table is fixed to a uniform distribution, so $P(\bar{F}) = P(F^-) = P(F^+) = \frac{1}{3}$. Thus, this class node is linked to the two other nodes D (discrete node) and X (Gaussian node), representing the GMM (see figure 6). Thus, we obtain the network represented on the figure 8, where we have also represented the different probability tables.

On the figure 8, we can make an interesting remark. The fact to have a GMM for one class of the system (here for the class \bar{F}) constrains the network to have a GMM of the same dimension for each class of the system. But for the two other classes (F^- and F^+), we have only one Gaussian, and not a GMM. In order to respect these classes, we can see that in the probability table of D (table representing the coefficient of the GMM for each classes), the first coefficient is fixed to 1 and all the others are fixed to 0. In this manner, the GMM is fixed to only one Gaussian (GMM of dimension 1) for the classes F^- and F^+ . In the same way, in the X conditional probability table, we can see that we have to fix some Gaussian laws for the combinations like $\{F^-, b\}$ or $\{F^+, c\}$ and so on. These Gaussians can be fixed randomly. Indeed, they will not be take into account in the inference, as their coefficient in the GMM is 0 for each.

We will evaluate the effectiveness of the proposed approach in the next section.

V. APPLICATION

We will apply the proposed method in order to prove his effectiveness. For that, we test our method on a simulated signal. Practically, this signal could represent a quality characteristic of a product to monitor. We have simulated 100 observations

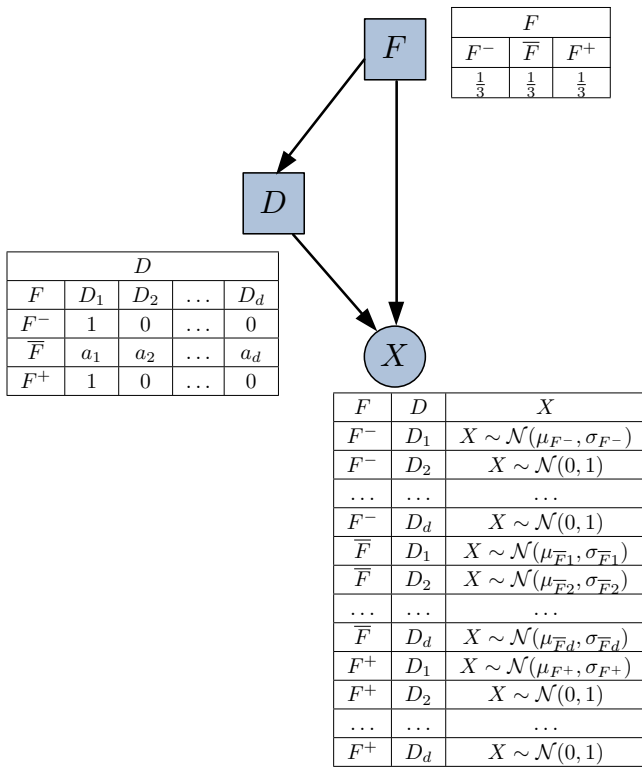


Fig. 8. BN for fault detection in Gaussian signal

of the normal operating conditions of this signal. The signal (in the fault-free case) can be seen on figure 9. We will see that is non-Gaussian cause it has some peaks at high values frequently.

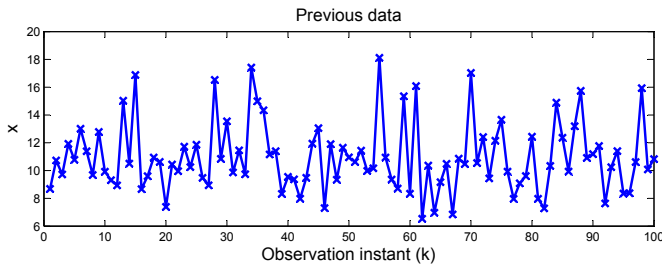


Fig. 9. Non-Gaussian signal in the fault-free case

We have applied the procedure proposed in the previous section to the data of figure 9. The false alarm risk α has been fixed to 0.1%; and we have made the assumption that this risk is equivalent on each side (so $\alpha/2$ on each side). The Bayesian network of figure 10 has been found.

To prove the effectiveness of the network, we have simulated 200 samples (observations) of this signal under the following scenario (see also figure 11):

- Samples 1 to 50: the system is in a fault-free case.
- Samples 51 to 100: a negative shift is assumed to occur after sample $k=51$ and to vanish after sample $k=100$.

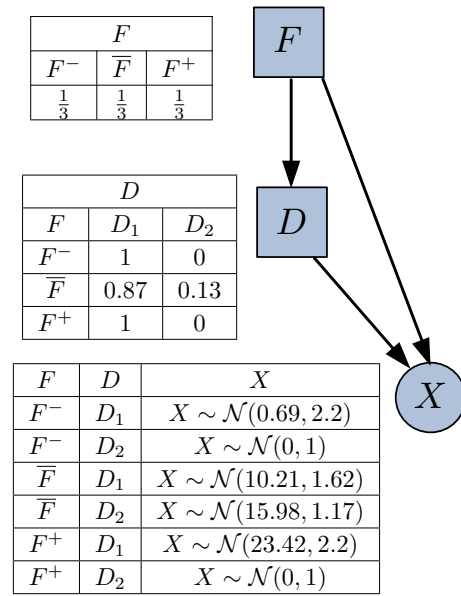


Fig. 10. Bayesian network of the application

- Samples 101 to 150: the system is in a fault-free case.
- Samples 151 to 200: a positive shift is assumed to occur after sample $k=151$ and to vanish after sample $k=200$.

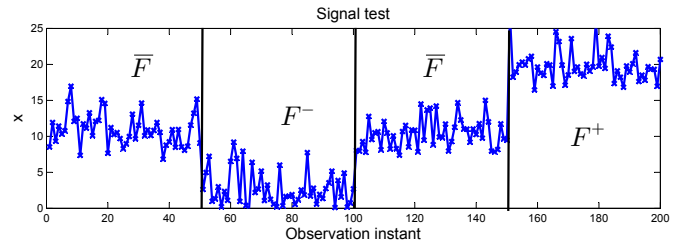


Fig. 11. Signal for the test of the network

We precise that all the simulations have been made with Matlab/Simulink and the BNT (BayesNet Toolbox) developed by Murphy ([27]).

The result of the simulation is given on the following graph (figure 12). This figure represents the a posteriori probabilities (so, given the observation x) of each class of the network. For a class, if the a posteriori probability is greater than 0.5, then we can say that the class defines the state of the signal.

On the figure 12, we see that each phase of the scenario is well recognized. We can view that we have no false alarm. This is normal because the false alarm risk chosen is quite high. We can also see that the fault F^- is better detected than the fault F^+ . The reason of this better detection is the fact that the distribution of F^- is more stretched on high value (as on figure 7). A non-symmetric α could be better to detect this kind of distribution. Finally, we can say that the proposed method performs well because for each scenario period, correct decisions are taken by the Bayesian network.

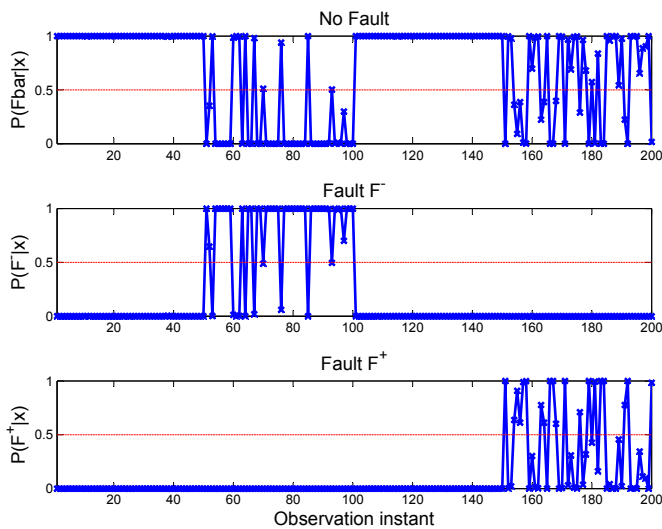


Fig. 12. Results of the simulation

VI. CONCLUSIONS AND OUTLOOKS

The main interest of this article is the presentation of a new method for the fault detection of industrial processes. This method is based on Bayesian networks using discrete and continuous (Gaussian) nodes. After the presentation of the classic case where the signal to monitor follows a Gaussian distribution, we have studied the case of a non-Gaussian signal. We have proposed a new strategy allowing to detect a change in a non-Gaussian signal. This strategy is based on the exploitation of Gaussian Mixture Model in the Bayesian network. The proposed method has been tested on a simple simulated non-Gaussian signal, where the constructed network gives good results.

The outlooks of this works are numerous. Indeed, we have developed several works in the field of process monitoring with Bayesian networks, but always in the Gaussian case. So, at present time, we will investigate the combination of the proposed method to previous works on Gaussian data.

REFERENCES

- [1] R. Isermann, *Fault Diagnosis Systems An Introduction from Fault Detection to Fault Tolerance*. Springer, 2006.
- [2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. Kavuri, "A review of process fault detection and diagnosis part i: Quantitative model-based methods," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.
- [3] R. J. Patton, P. M. Frank, and R. N. Clark, *Issues of Fault Diagnosis for Dynamic Systems*. Springer, 2000.
- [4] R. Isermann and P. Ball, "Trends in the application of model based fault detection and diagnosis of technical processes," in *World IFAC Congress*, San Francisco, USA, 1996, pp. 1–12.
- [5] P. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. a survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [6] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, and K. Yin, "A review of process fault detection and diagnosis part iii: Process history based methods," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003.
- [7] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault detection and diagnosis in industrial systems*. New York: Springer-Verlag, 2001.

- [8] T. Denoeux, M. Masson, and B. Dubuisson, "Advanced pattern recognition techniques for system monitoring and diagnosis : A survey," *Journal Europeen des Systemes Automatises*, vol. 31, no. 9-10, pp. 1509–1539, 1997.
- [9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification 2nd edition*. Wiley, 2001.
- [10] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.
- [11] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, pp. 21–27, 1967.
- [12] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.
- [13] N. Mehranbod, M. Soroush, M. Piovoso, and B. Ogunnaike, "Probabilistic model for sensor fault detection and identification," *AIChE Journal*, vol. 49, no. 7, pp. 1787–1802, 2003.
- [14] G. Weidl, A. Madsen, and S. Israelson, "Applications of object-oriented bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes," *Computers and Chemical Engineering*, vol. 29, no. 9, pp. 1996–2009, 2005.
- [15] S. Verron, T. Tiplica, and A. Kobi, "Procedure based on mutual information and bayesian networks for fault diagnosis of industrial systems," in *American Control Conference*, 2007, pp. 420–425.
- [16] —, "Distance rejection in a bayesian network for fault diagnosis of industrial systems," in *16th Mediterranean Conference on Control and Automation, MED'08*, Ajaccio, France, 2008.
- [17] K. P. Murphy, "Dynamic bayesian networks: Representation, inference and learning," Ph.D. dissertation, U.C. Berkeley, 2002.
- [18] P. Weber and M.-C. Suhner, "Modlisation de processus industriels par reseaux bayesiens orientes objet (rboo)," *Revue d'Intelligence Artificielle*, vol. 18, pp. 299–326, 2004.
- [19] F. V. Jensen, *An introduction to Bayesian Networks*. Taylor and Francis, London, United Kingdom, 1996.
- [20] S. Verron, T. Tiplica, and A. Kobi, "Multivariate control charts with a bayesian network," in *4th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2007, pp. 228–233.
- [21] B. Cobb, R. Rumi, and A. Salmeron, "Modeling conditional distributions of continuous variables in bayesian networks," in *Lecture Notes in Computer Science*, vol. 3646 LNCS, 2005, pp. 36–45.
- [22] Y. Gurwicz and B. Lerner, "Rapid spline-based kernel density estimation for bayesian networks," in *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 700–703.
- [23] G. McLachlan and K. Basford, *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, 1988.
- [24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. B 39, pp. 1–38, 1977.
- [25] J. Liu, F. Liang, and W. Wong, "The use of multiple-try method and local optimization in metropolis sampling," *Journal of the American Statistical Association*, vol. 95, pp. 121–134, 2000.
- [26] S. Vempala and G. Wang, "A spectral algorithm for learning mixture models," *Journal of Computer and System Sciences*, vol. 68, no. 4, pp. 841–860, 2004.
- [27] K. P. Murphy, "The bayes net toolbox for matlab," in *In Computing Science and Statistics : Proceedings of Interface*, 2001.