Monitoring of complex processes with Bayesian networks

Sylvain Verron, Teodor Tiplica, Abdessamad Kobi

0

Monitoring of complex processes with Bayesian networks

Sylvain Verron, Teodor Tiplica and Abdessamad Kobi
LASQUO/ISTIA - University of Angers
France

1. Introduction

Industrial processes are more and more complex and include a lot of sensors giving measurements of some attributes of the system. A study of these measurements can allow to decide on the correct working conditions of the process. If the process is not in normal working conditions, it signifies that a fault has occurred in the process. If no fault has occurred, thus the process is in the fault-free case. An important research field is on the Fault Detection and Diagnosis (FDD) (Isermann (2006)). The goal of a FDD scheme is to detect, the earliest possible, when a fault occurs in the process. Once the fault has been detected, the other important step is the diagnosis. The diagnosis can be seen as the decision of which fault has appeared in the process, what are the characteristics of this fault, what are the root causes of the fault.
One can distinguish three principal categories of methods for the FDD (Chiang et al. (2001)): the knowledge-based approach, the model-based approach and the data-driven approach. The knowledge-based category represents methods based on qualitative models (FMECA - Failures Modes Effects and Critically Analysis; Fault Trees; Decision Trees; Risk Analysis) (Stamatis (2003); Dhillon (2005)). For the model-based methods, an analytical model of the process is constructed based on the physical relations governing the process (Patton et al. (2000)). The model gives the normal (fault free) value of each sensor or variable of the system for each sample instant, then residuals are generated (residuals are the differences between measurements and the corresponding reference values estimated with the model of the fault-free system). If the system is fault free, residuals are almost nil, and so their evaluations allow to detect and diagnose a fault. Theoretically, the best methods are the analytical ones, but the major drawback of this family of techniques is the fact that a detailed model of the process is required in order to monitor it efficiently. Obtaining an effective detailed model can be very difficult, time consuming and expensive, particularly for large-scale systems with many variables. The last category of methods are the process history (or data-driven) methods (Venkatasubramanian et al. (2003)). These techniques are based on rigorous statistical developments of process data. In literature, we can find many different data-driven techniques for FDD. For the fault detection of industrial processes many methods have been submitted: univariate statistical process control (Shewhart charts) (Montgomery (1997)), multivariate statistical process control (T^2 and Q charts) (Westerhuis et al. (2000)), and some Principal Component Analysis (PCA) based techniques (Jackson (1985)). Kano et al. (2002) make comparisons between these different techniques. For the fault diagnosis techniques we can cite the book of Chiang et al. (2001) which presents a lot of them (PCA based techniques, Fisher Discriminant Analysis, PLS based techniques, etc).

The purpose of this article is to present application of a promising tool for the Fault Detection and Diagnosis: the Bayesian network. The aim of the paper is to demonstrate that some FDD techniques can be modeled very simply in a Bayesian network, with very good performances. The article is structured in the following manner. In section 2, we introduce different notions (theoretical and practical) about Bayesian network. The section 3 presents how to model multivariate control charts in a Bayesian network, in order to make an effective way for the fault detection by the Bayesian network. In the same way, section 4 presents the modeling of discriminant analysis by Bayesian network for fault diagnosis of systems. The section 5 presents an evaluation of the proposed method for detection and diagnosis of faults on the benchmark Tennessee Eastman Problem. Finally, we conclude on this method and present some perspectives.

2. Bayesian network

A Bayesian Network (BN) (Pearl (1988)) is a probabilistic graphical model where each variable is a node. Edges of the graph represent dependences between linked nodes. A formal definition of Bayesian network (Jensen (1996)) is a couple $\{\mathbf{G}, \mathbf{P}\}$ where:

$\{\mathbf{G}\}$ is a directed acyclic graph, whose nodes are random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ and whose missing edges represent conditional independences between the variables,

$\{\mathbf{P}\}$ is a set of conditional probability distributions (one for each variable): $P = \{p(X_1|pa(X_1)), \ldots, p(X_n|pa(X_n))\}$ where $p(X_i|pa(X_i))$ is a table defined by $p(X_i = x_i^j|pa(X_i))$ with $x_i^j \in Dom(X_i) = x_i^1, x_i^2, ..., x_i^{n_i}$ where $Dom(X_i)$ is the set of modalities of variable X_i and n_i is the number of these modalities. The joint probability should read like the following equation:

$$p(x) = \prod_{i=1}^{n} (X_i|pa(X_i)) \tag{1}$$

with $x = (x_1^{j_1}, x_2^{j_2}, ..., x_n^{j_n})$.

Theoretically, variables $X_1, X_2, \ldots, X_n$ can be discrete or continuous. However, in practice, for exact computation, only the discrete and the Gaussian case can be treated. Such a network is often called Conditional Gaussian Network (CGN). In this context, to ensure availability of exact computation methods, discrete variables are not allowed to have continuous parents (see Lauritzen & Jensen (2001); Madsen (2008)).

In concrete terms, the conditional probability distribution is described for each node by his Conditional Probability Table (CPT). In a CGN, three cases of CPT can be found. The first one is for a discrete variable with discrete parents. For example, we take the case of two discrete variables A and B of respective dimensions a and b (with $a_1, a_2, \ldots, a_a$ the different modalities of A, and $b_1, b_2, \ldots, b_b$ the different modalities of B). If A is parent of B, then the CPT of B is represented in table 1.

As we can see, the utility of the CPT is to condense the information about the relations of B with his parents. We can denote that the dimension of this CPT (number of conditional probabilities) is $a \times b$. In general the dimension of the CPT of a discrete node (dimension a) with p parents (discrete) $Y_1, Y_2, \ldots, Y_p$ (dimension $y_1, y_2, \ldots, y_p$) is $a \times \prod_{i=1}^{p} y_i$.

The second case of CPT is for a continuous variable with discrete parents. Assuming that B is a Gaussian variable, and that A is a discrete parent of B with a modalities, the CPT of B can

	B			
A	b_1	b_2	$\dots$	b_b
a_1	$P(b_1\|a_1)$	$P(b_2\|a_1)$	$\dots$	$P(b_b\|a_1)$
a_2	$P(b_1\|a_2)$	$P(b_2\|a_2)$	$\dots$	$P(b_b\|a_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_a	$P(b_1\|a_a)$	$P(b_2\|a_a)$	$\dots$	$P(b_b\|a_a)$

Table 1. CPT of a discrete node with discrete parents

be represented as in the table 2 where $P(B|a_1) \sim \mathcal{N}(\mu_{a_1}, \Sigma_{a_1})$ indicates that B conditioned to $A = a_i$ follows a multivariate normal density function with parameters μ_{a_i} and Σ_{a_i}.

A	B
a_1	$P(B\|a_1) \sim \mathcal{N}(\mu_{a_1}, \Sigma_{a_1})$
a_2	$P(B\|a_2) \sim \mathcal{N}(\mu_{a_2}, \Sigma_{a_2})$
$\vdots$	$\vdots$
a_a	$P(B\|a_a) \sim \mathcal{N}(\mu_{a_a}, \Sigma_{a_a})$

Table 2. CPT of a Gaussian node with discrete parents

The third case occurs when a continuous node B has a continuous parent A. In this case, we obtain a linear regression and we can write, for a fixed value a of A, that B follows a Gaussian distribution $P(B|A = a) \sim \mathcal{N}(\mu_B + \beta \times a; \Sigma_B)$ where β is the regression coefficient. The three different cases of CPT enumerated can evidently be combined for different cases where a continuous variable has several discrete parents and several continuous (Gaussian) parents.

The classical use of a Bayesian network (or Conditional Gaussian Network) is to enter evidence in the network (an evidence is the observation of the values of a set of variables). Therefore, the information given by the evidence is propagated in the network in order to update the knowledge and obtain a posteriori probabilities on the non-observed variables. This propagation mechanism is called inference. As its name suggests, in a Bayesian network, the inference is based on the Bayes rule. A lot of inference algorithms (exact or approximate) have been developed, but one of the more exploited is the junction tree algorithm (Jensen et al. (1990)).

Bayesian network classifiers are particular BN 5Friedman et al. (1997). They always have a discrete node C coding the k different classes of the system. Thus, other variables $X_1, \dots, X_p$ represent the p descriptors (variables) of the system.

A famous Bayesian classifier is the Naïve Bayesian Network (NBN), also named Bayes classifier (Langley et al. (1992)). This Bayesian classifier makes the strong assumption that the descriptors of the system are class conditionally independent. Assuming the hypothesis of normality of each descriptor, the NBN is equivalent to the classification rule of the diagonal quadratic discriminant analysis. But, in practice, this assumption of independence and non-correlated variables is not realistic. In order to deal with correlated variables, several approaches have been developed. We can cite the Tree Augmented Naïve Bayesian networks (TAN) (Friedman et al. (1997)). These BNs are based on a NBN but a tree is added between the descriptors. An other interesting approach is the Kononenko one (Kononenko (1991)), which represents some variables in one node. As in (Perez et al. (2006)) the assumption we will make is that this variable follows a normal multivariate distribution (conditionally to the class) and we will refer to this kind of BN as Condensed Semi Naïve Bayesian Network (CSNBN).

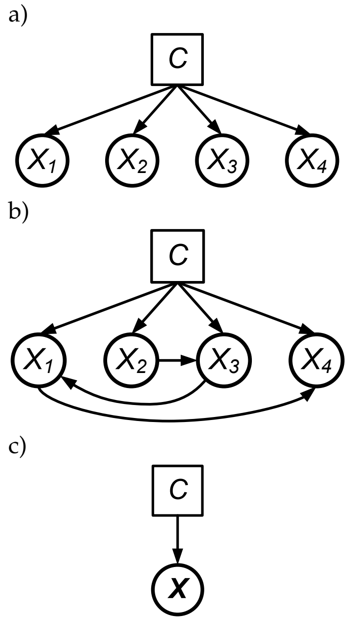


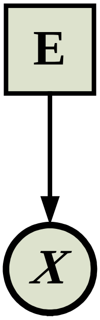
Fig. 1. Different bayesian network classifiers: NBN (a), TAN (b) and CSNBN (c).

3. Fault detection with Bayesian network

In previous work (Verron et al. (2007b)), we have demonstrated that a T^2 control chart Hotelling (1947) could be modeled with a Bayesian network. For that, we use two nodes: a Gaussian multivariate node $\boldsymbol{X}$ 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 Σ are respectively the mean vector and the variance-covariance matrix of the process, we can monitor the process with the following rule: if $P(IC|\boldsymbol{x}) < P(IC)$ then the process is out-of-control. This Bayesian network is represented on the Figure 2, where the conditional probabilities tables of each node are given.
In Figure 2, we can observe that a coefficient c is implicated in the modeling of the control chart by Bayesian network. This coefficient is the root (different of 1) of the following equation:

$$1 - c + \frac{pc}{CL}\ln(c) = 0 \tag{2}$$

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. The demonstration of the computation of c is given in A. In numerous cases, CL is equal to $\chi^2_{\alpha,p}$, the quantile at the value α of the distribution of the χ^2 with p degrees of freedom (Montgomery (1997)). α allows us to tune the false alarm rate of the control chart.



E	
IC	OC
$P(IC)$	$P(OC)$

E	$\boldsymbol{X}$
IC	$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$
OC	$\boldsymbol{X} \sim N(\boldsymbol{\mu}, c \times \Sigma)$

Fig. 2. T^2 control chart in a Bayesian network

The application of this network to a two variables process is given in figure 3.

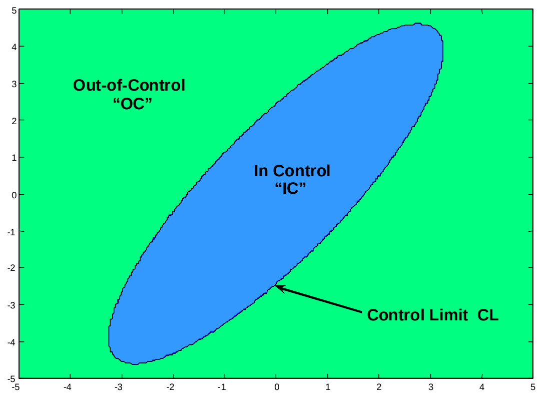
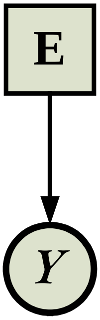


Fig. 3. Detection area of the Bayesian network

A particular interest of the modeling of control chart in a Bayesian network is that a MEWMA control chart (Lowry et al. (1992)) can also be modeled in the same way. The principle of the MEWMA control chart is to take into account the process evolution in weighting past observations extracted from the process. The MEWMA variable $\boldsymbol{y}_t$ is computed recursively, for each sample, by the equation 3 where the initialization is given by $\boldsymbol{y}_0 = \boldsymbol{\mu}$.

$$\boldsymbol{y}_t = \lambda \boldsymbol{x}_t + (1-\lambda)\boldsymbol{y}_{t-1} \tag{3}$$

In the same way that the T^2 control chart, we can also monitor the process with a MEWMA control chart modeled by the Bayesian network of the figure 2.
We can precise that performances of the MEWMA control chart are function of λ. Indeed, a small λ allows a performing detection of small magnitude shifts, but a higher λ will be more adapted for large magnitude shifts. So, the choice of λ will be function of the magnitude shift that one wants to detect. A particular case of the MEWMA control chart is the case where $\lambda = 1$. In this case, the MEWMA chart is equivalent to the T^2 control chart.



E	
IC	OC
$P(IC)$	$P(OC)$

E	$\boldsymbol{Y}$
IC	$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \left(\frac{\lambda}{2-\lambda}\right)\Sigma)$
OC	$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, c \times \left(\frac{\lambda}{2-\lambda}\right)\Sigma)$

Fig. 4. MEWMA control chart in a Bayesian network

4. Bayesian network for fault diagnosis

Once a problem (fault) has been detected in the evolution of the process by the mean of a detection method, we need to identify (diagnosis) the belonging class of this fault. Thereby, the diagnosis problem can be viewed as the task to correctly classify this fault in one of the predefined fault classes. The classification task needs the construction of a classifier (a function allocating a class to the observations described by the variables of the system). Two types of classification exist: unsupervised classification which objective is to identify the number and the composition of each class present in the data structure; supervised classification where the number of classes and the belonging class of each observation is known in a learning sample and whose objective is to class new observations to one of the existing classes. For example, given a learning sample of a bivariate system with three different known faults as illustrated in the figure 5, we can easily use supervised classification to classify a new faulty observation. A feature selection can be used in order to select only the most informative variables of the problem (Verron et al. (2008)). In this study, we will use the Bayesian network as a supervised classification tool.

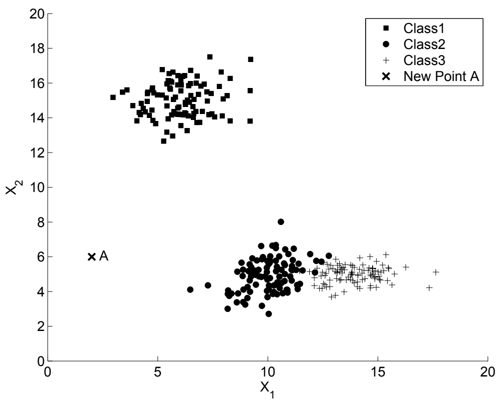
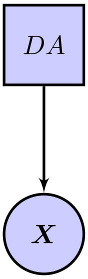


Fig. 5. Bivariate system with three different known faults

In the context of the diagnosis of industrial systems, Bayesian networks and Conditional Gaussian Networks have been already used and they give convenient results compared to other classification tools like support vector machines, neural networks or k-nearest neighborhoods (Pernkopf (2005); Perzyk et al. (2005); Tiplica et al. (2006); Verron et al. (2007a;c)). As the performances of the CGN have been previously demonstrated (Verron et al. (2007a;c)), we choose this classifier in this article which is equivalent to a Discriminant Analysis (DA). Therefore, we name the class node DA, and the observation node $\boldsymbol{X}$ (a normal multivariate node). The figure 6 presents the CGN equivalent to a discriminant analysis, with the probability tables associated to each node. To simplify, the a priori probability of each class F_i is fixed to $p(F_i) = \frac{1}{k}$, where k is the number of known faults. The node $\boldsymbol{X}$ follows the different normal probability densities ($\mathcal{N}$) conditionally to the class of DA, where $\boldsymbol{\mu}_i$ is the mean vector of the fault F_i, Σ_i is the covariance matrix of the fault F_i. $\boldsymbol{\mu}_i$ and Σ_i are estimated on the fault database by Maximum Likelihood Estimation (MLE) (Duda et al. (2001)). In the mere example of the figure 5, the CGN gives the different areas of classification of the figure 7.



DA		
F_1	$\dots$	F_k
$\frac{1}{k}$	$\dots$	$\frac{1}{k}$

DA	$\boldsymbol{X}$
F_1	$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1)$
$\dots$	$\dots$
F_k	$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k)$

Fig. 6. Conditional Gaussian Network equivalent to a discriminant analysis

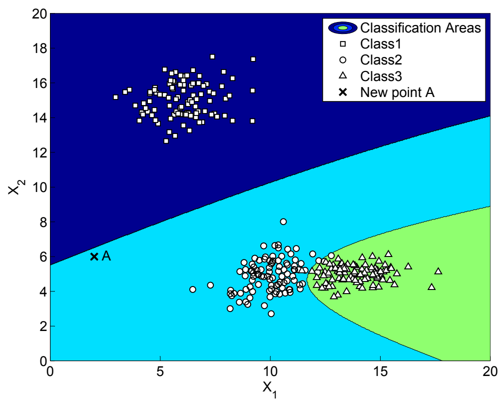


Fig. 7. Classification areas of the bivariate system

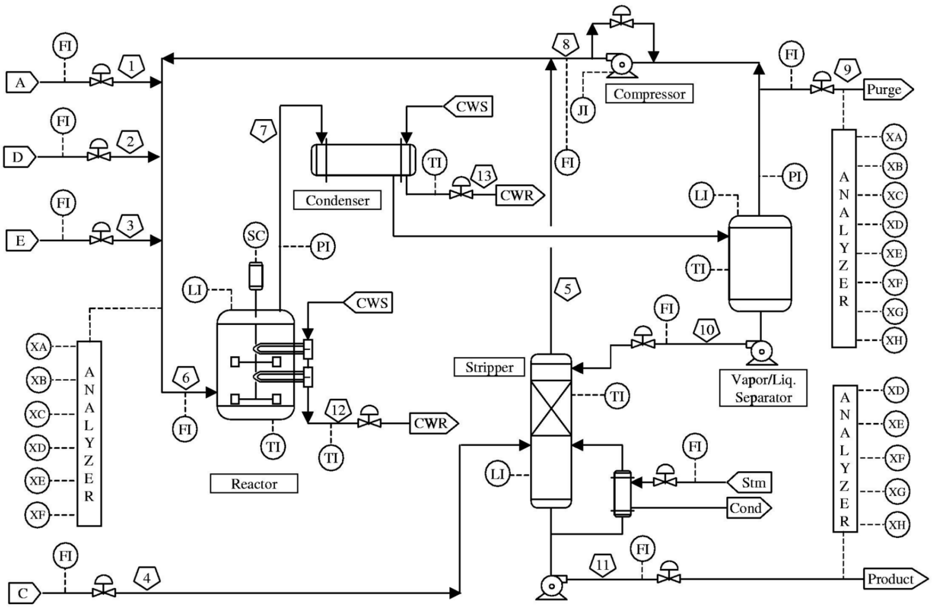


Fig. 8. Process flowsheet of the TEP

5. Application to the TEP

Now, we are going to study an application of the Bayesian network approach on a benchmark problem: the Tennessee Eastman Process (figure 8).

5.1 Presentation of the TEP

We have tested our approach on the Tennessee Eastman Process. The Tennessee Eastman Process (TEP) is a chemical process. It is not a real process but a simulation of a process that was created by the Eastman Chemical Company to provide a realistic industrial process in order to evaluate process control and monitoring methods. The article of Downs & Vogel (1993) entirely describes this process. The authors also give the Fortran code of the simulation of the process. Ricker (1996) has implemented the simulation on Matlab. The TEP is composed of five major operation units: a reactor, a condenser, a compressor, a stripper and a separator. Four gaseous reactants A, C, D, E and an inert one B are fed to the reactor where the liquid products F, G and H are formed. This process has 12 input variables and 41 output variables. The TEP has 20 types of identified faults. This process is ideal to test monitoring methods. However, it is also a benchmark problem for control techniques because it is open-loop unstable. A lot of articles present the TEP and test their approaches on it. For example, in fault detection, we can cite Kano et al. (2002) and Kruger et al. (2004). Some fault diagnosis techniques have also been tested on the TEP (Chiang et al. (2001; 2004); Kulkarni et al. (2005); Maurya et al. (2007)) with the plant-wide control structure recommended in Lyman & Georgakis (1995).
As indicated in the table 3, each type of fault is composed of 2 datasets: a training sample and a testing sample, containing respectively 480 and 800 observations. We precise that in the

next part of this paper all computations have been made on Matlab with the BNT (BayesNet Toolbox) developed by Murphy (2001).

Class	Train data	Test data
Fault free	480	800
Fault 1	480	800
Fault 2	480	800
...	...	...
Fault k	480	800
...	...	...
Fault 20	480	800

Table 3. Data of the TEP

5.2 Detection

In order to test the performances of the Bayesian network approach for the detection, we set an acceptable false alarm for the detection of 0.01 (1%). As the detection is modeled with two control chart, the local false alarm rate is set to 0.005. The table 4 presents the results of the Bayesian network dedicated to the detection, composed of the modeling of the T^2 and MEWMA control charts.

Fault	First detection instant	Detection rate
1	3	99.75
2	13	98.5
3	34	35
4	1	100
5	1	100
6	1	100
7	1	100
8	18	97.75
9	7	15.88
10	18	97
11	7	90.88
12	2	99.88
13	37	95.5
14	1	100
15	146	30.5
16	9	99
17	20	97.5
18	57	92.38
19	2	96.5
20	65	91.88
Mean	22.15	91.38

Table 4. Detection results

In table 4, we can affirm that faults F3, F9 and F15 are very difficult to detect. Chiang et al. (Chiang et al. (2004)), using PCA (Principal Component Analysis) based method, on the same data, have made the same conclusions on these 3 faults. However, without these 3 faults, the mean detection rate of the other faults is more than 97.44% and proves the efficiency of the Bayesian network for the fault detection task.

5.3 Diagnosis

Always on the same data, we have applied the method proposed in section 4. After the learning of the parameters of the Bayesian network, we have presented 16 000 observations to the network (800 observations of each 20 faults). The network has given probabilities of each observation to come from each known faults. The decision of the fault has been taken for the fault with the maximum a posteriori probability. Results of the 16 000 observations are given in the table 6 of appendix B. A more readable table of results is given in table 5.

Fault	Diagnosis rate
1	97,5
2	98,12
3	22
4	82,37
5	98
6	100
7	100
8	97
9	22,62
10	86,87
11	75,5
12	98,25
13	76,12
14	98,75
15	23,5
16	80,62
17	85
18	68,5
19	96,12
20	87,37
Mean	79.71

Table 5. Diagnosis results

In the table 5, we can observe that, like for the fault detection, the faults F3, F9 and F15 are difficult to diagnose. Indeed, these three faults are very similar to the fault free conditions, and so they are difficult to detect and difficult to diagnose. However, for the other faults, we can notice that a lot of observations are correctly classified, and without the 3 difficult faults (F3, F9 and F15), the mean diagnosis rate increase to 90%.

6. Conclusion

In this chapter, we have studied the application of Bayesian networks (and more particularly of Conditional Gaussian networks) for the fault detection and diagnosis. The fault detection is made by a modeling of multivariate control charts (T^2 and MEWMA) with Bayesian network. On the same way, the fault diagnosis is similar to a supervised classification task. A Bayesian network is able to discriminate between different faults of a system. For that, we have modeled a discriminant analysis directly in the Bayesian network. The performances of the proposed approach are evaluated on the benchmark problem of the Tennessee Eastman Process, demonstrating that fault detection and fault diagnosis can be made with Bayesian network. Outlooks of this work are on the use a Bayesian network as a causal model of a process, in order to realize fault isolation of the different variables implicated in a fault.

A. Coefficient c demonstration

This appendix presents the demonstration of the equation 2.
As in the case of the T^2 control chart (Montgomery (1997)), we will fix a threshold (Control Limit CL for the control chart) on the a posteriori probabilities allowing to take decisions on the process: if, for a given observation $\boldsymbol{x}$, the a posteriori probability to be allocated to F_i ($P(F_i|\boldsymbol{x})$) is greater than the a priori probability to be allocated to F_i ($P(F_i)$), then this observation is allocated to F_i. This rule can be rewritten as: $\boldsymbol{x} \in F_i$ if $P(F_i|\boldsymbol{x}) > P(F_i)$, or equivalently $\boldsymbol{x} \in \overline{F_i}$ if $P(\overline{F_i}|\boldsymbol{x}) < P(\overline{F_i})$. The objective of the following developments is to define c in order to obtain the equivalence between the CGN and the multivariate T^2 control chart.
We want to keep the following decision rule:

$$\boldsymbol{x} \in F_i \quad if \quad T^2 < CL \tag{4}$$

with this decision rule:

$$\boldsymbol{x} \in F_i \quad if \quad P(F_i|\boldsymbol{x}) > P(F_i) \tag{5}$$

We develop the second decision rule:

$$\begin{aligned}
P(F_i|\boldsymbol{x}) &> P(F_i) \\
P(F_i|\boldsymbol{x}) &> (P(F_i))(P(F_i|\boldsymbol{x}) + P(\overline{F_i}|\boldsymbol{x})) \\
P(F_i|\boldsymbol{x}) &> P(F_i)P(F_i|\boldsymbol{x}) + P(F_i)P(\overline{F_i}|\boldsymbol{x}) \\
P(F_i|\boldsymbol{x}) - P(F_i)P(F_i|\boldsymbol{x}) &> P(F_i)P(\overline{F_i}|\boldsymbol{x}) \\
P(F_i|\boldsymbol{x})(1 - P(F_i) &> P(F_i)P(\overline{F_i}|\boldsymbol{x}) \\
P(F_i|\boldsymbol{x})P(\overline{F_i}) &> P(F_i)P(\overline{F_i}|\boldsymbol{x}) \\
P(F_i|\boldsymbol{x}) &> \frac{P(F_i)}{P(\overline{F_i})}P(\overline{F_i}|\boldsymbol{x})
\end{aligned}$$

However, the Bayes law gives:

$$P(F_i|\boldsymbol{x}) = \frac{P(F_i)P(\boldsymbol{x}|F_i)}{P(\boldsymbol{x})} \tag{6}$$

and

$$P(\overline{F_i}|\boldsymbol{x}) = \frac{P(\overline{F_i})P(\boldsymbol{x}|\overline{F_i})}{P(\boldsymbol{x})} \tag{7}$$

As a consequence, we obtain:

$$\begin{aligned}
\frac{P(F_i)P(\boldsymbol{x}|F_i)}{P(\boldsymbol{x})} &> (\frac{P(F_i)}{P(\overline{F_i})})\frac{P(\overline{F_i})P(\boldsymbol{x}|\overline{F_i})}{P(\boldsymbol{x})} \\
(\frac{P(F_i)}{P(\overline{F_i})})P(\boldsymbol{x}|F_i) &> (\frac{P(F_i)}{P(\overline{F_i})})P(\boldsymbol{x}|\overline{F_i}) \\
P(\boldsymbol{x}|F_i) &> P(\boldsymbol{x}|\overline{F_i}) \qquad (8)
\end{aligned}$$

In the case of a discriminant analysis with k classes C_i, the conditional probabilities are computed with the following equation 9, where ϕ represents the probability density function of the multivariate Gaussian distribution of the class.

$$P(\boldsymbol{x}|C_i) = \frac{\phi(\boldsymbol{x}|C_i)}{\sum_{j=1}^{k} P(C_j)\phi(\boldsymbol{x}|C_j)} \qquad (9)$$

Equation 8 can be written as:

$$\phi(\boldsymbol{x}|F_i) > \phi(\boldsymbol{x}|\overline{F_i}) \qquad (10)$$

We recall that the probability density function of a multivariate Gaussian distribution of dimension p, of parameters $\boldsymbol{\mu}$ and Σ, of an observation $\boldsymbol{x}$ is given by:

$$\phi(\boldsymbol{x}) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{(2\pi)^{p/2}|\Sigma|^{1/2}} \qquad (11)$$

If the law parameters are $\boldsymbol{\mu}$ and $c \times \Sigma$, then the density function becomes:

$$\phi(\boldsymbol{x}) = \frac{e^{-\frac{1}{2c}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{(2\pi)^{p/2}|\Sigma|^{1/2}c^{p/2}} \qquad (12)$$

In identifying the expression $(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ as the T^2 of the observation $\boldsymbol{x}$, we can write:

$$\begin{aligned}
\phi(\boldsymbol{x}|F_i) &> \phi(\boldsymbol{x}|\overline{F_i}) \\
\frac{e^{-\frac{T^2}{2}}}{(2\pi)^{p/2}|\Sigma|^{1/2}} &> \frac{e^{-\frac{T^2}{2c}}}{(2\pi)^{p/2}|\Sigma|^{1/2}c^{p/2}} \\
e^{-\frac{T^2}{2}} &> \frac{e^{-\frac{T^2}{2c}}}{c^{p/2}} \\
-\frac{T^2}{2} &> -\frac{T^2}{2c} - \frac{p\ln(c)}{2} \\
T^2 &< \frac{T^2}{c} + p\ln(c) \\
T^2 &< \frac{p\ln(c)}{1-\frac{1}{c}} \qquad (13)
\end{aligned}$$

However, we search the value(s) of c allowing the equivalence with the control chart decision rule: $\boldsymbol{x} \in F_i \quad if \quad T^2 < CL$. So, we obtain the following equation for c:

$$\frac{p \ln(c)}{1 - \frac{1}{c}} = LC \tag{14}$$

Or, equivalently:

$$1 - c + \frac{pc}{LC} \ln(c) = 0 \tag{15}$$

B. Fault diagnosis detailed results

!

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
F1	780	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
F2	0	785	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F3	0	0	176	0	0	0	0	2	201	8	18	0	16	0	118	15	1	38	6	17
F4	0	0	0	659	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0
F5	0	0	0	0	784	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
F6	0	0	0	0	0	800	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F7	0	0	0	0	0	0	800	0	0	0	0	0	0	0	0	0	0	0	0	0
F8	18	5	1	0	0	0	0	776	0	4	0	1	109	0	1	26	0	0	0	0
F9	0	8	171	0	0	0	0	11	181	25	24	1	4	0	233	15	7	13	9	34
F10	0	0	48	0	0	0	0	0	40	695	9	0	0	0	64	48	0	3	5	6
F11	0	0	43	141	0	0	0	3	42	5	604	0	2	1	43	2	30	3	2	3
F12	0	0	0	0	16	0	0	4	0	6	0	786	41	0	4	10	0	168	0	23
F13	0	0	0	0	0	0	0	2	0	3	0	1	609	0	3	4	0	0	0	3
F14	0	0	17	0	0	0	0	0	10	3	28	0	0	790	20	4	71	0	0	1
F15	0	1	215	0	0	0	0	0	221	12	34	0	11	0	188	6	9	9	2	7
F16	0	1	85	0	0	0	0	0	39	35	5	0	2	0	82	645	1	10	4	3
F17	1	0	7	0	0	0	0	0	6	3	42	0	0	9	1	3	680	0	1	2
F18	0	0	0	0	0	0	0	0	0	0	0	8	5	0	0	0	0	548	1	0
F19	1	0	32	0	0	0	0	0	54	1	7	0	1	0	38	16	1	1	769	2
F20	0	0	5	0	0	0	0	0	6	0	2	0	0	0	5	6	0	7	1	699

Table 6. Confusion matrix

C. References

Chiang, L. H., Russell, E. L. & Braatz, R. D. (2001). *Fault detection and diagnosis in industrial systems*, New York: Springer-Verlag.

Chiang, L., Kotanchek, M. & Kordon, A. (2004). Fault diagnosis based on fisher discriminant analysis and support vector machines, *Computers and Chemical Engineering* **28**(8): 1389–1401.

Dhillon, B. (2005). *Reliability,Quality,and Safety for Engineers*, CRC Press.

Downs, J. & Vogel, E. (1993). Plant-wide industrial process control problem, *Computers and Chemical Engineering* **17**(3): 245–255.

Duda, R. O., Hart, P. E. & Stork, D. G. (2001). *Pattern Classification 2nd edition*, Wiley.

Friedman, N., Geiger, D. & Goldszmidt, M. (1997). Bayesian network classifiers, *Machine Learning* **29**(2-3): 131–163.

Hotelling, H. (1947). Multivariate quality control, *Techniques of Statistical Analysis* : 111–184.

Isermann, R. (2006). *Fault Diagnosis Systems An Introduction from Fault Detection to Fault Tolerance*, Springer.

Jackson, E. J. (1985). Multivariate quality control, *Communication Statistics - Theory and Methods* **14**: 2657 – 2688.

Jensen, F. V. (1996). *An introduction to Bayesian Networks*, Taylor and Francis, London, United Kingdom.

Jensen, F. V., Lauritzen, S. L. & Olesen, K. G. (1990). Bayesian updating in causal probabilistic networks by local computations, *Computational Statistics Quaterly* **4**: 269–282.

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

Kononenko, I. (1991). Semi-naive bayesian classifier, *EWSL-91: Proceedings of the European working session on learning on Machine learning*, pp. 206–219.

Kruger, U., Zhou, Y. & Irwin, G. (2004). Improved principal component monitoring of large-scale processes, *Journal of Process Control* **14**(8): 879–888.

Kulkarni, A., Jayaraman, V. & Kulkarni, B. (2005). Knowledge incorporated support vector machines to detect faults in tennessee eastman process, *Computers and Chemical Engineering* **29**(10): 2128–2133.

Langley, P., Iba, W. & Thompson, K. (1992). An analysis of bayesian classifiers, *National Conference on Artificial Intelligence*.

Lauritzen, S. & Jensen, F. (2001). Stable local computation with conditional gaussian distributions, *Statistics and Computing* **11**(2): 191–203.

Lowry, C. A., Woodall, W. H., Champ, C. W. & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart, *Technometrics* **34**(1): 46–53.

Lyman, P. & Georgakis, C. (1995). Plant-wide control of the tennessee eastman problem, *Computers and Chemical Engineering* **19**(3): 321–331.

Madsen, A. (2008). Belief update in clg bayesian networks with lazy propagation, *International Journal of Approximate Reasoning* **49**(2): 503–521.

Maurya, M. R., Rengaswamy, R. & Venkatasubramanian, V. (2007). Fault diagnosis using dynamic trend analysis: A review and recent developments, *Engineering Applications of Artificial Intelligence* **20**(2): 133–146.

Montgomery, D. C. (1997). *Introduction to Statistical Quality Control, Third Edition*, John Wiley and Sons.

Murphy, K. P. (2001). The bayes net toolbox for matlab, *Computing Science and Statistics : Proceedings of Interface*, Vol. 33, pp. 33–40.

Patton, R. J., Frank, P. M. & Clark, R. N. (2000). *Issues of Fault Diagnosis for Dynamic Systems*, Springer.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers.

Perez, A., Larranaga, P. & Inza, I. (2006). Supervised classification with conditional gaussian networks: Increasing the structure complexity from naive bayes, *International Journal of Approximate Reasoning* **43**: 1–25.

Pernkopf, F. (2005). Bayesian network classifiers versus selective k-nn classifier, *Pattern Recognition* **38**(1): 1–10.

Perzyk, M., Biernacki, R. & Kochanski, A. (2005). Modeling of manufacturing processes by learning systems: The naive bayesian classifier versus artificial neural networks, *Journal of Materials Processing Technology* **164-165**: 1430–1435.

Ricker, N. (1996). Decentralized control of the tennessee eastman challenge process, *Journal of Process Control* **6**(4): 205–221.

Stamatis, D. H. (2003). *Failure Mode and Effect Analysis: FMEA from Theory to Execution*, ASQ Quality Press.

Tiplica, T., Verron, S., Kobi, A. & Nastac, I. (2006). Fdi in multivariate process with naïve bayesian network in the space of discriminant factors, *International Conference on Computational Intelligence for Modelling, Control and Automation*, Sydney, Australia, p. 216.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. & Yin, K. (2003). A review of process fault detection and diagnosis part iii: Process history based methods, *Computers and Chemical Engineering* **27**(3): 327–346.

Verron, S., Tiplica, T. & Kobi, A. (2007a). Fault diagnosis of industrial systems with bayesian networks and mutual information, *European Control Conference*, pp. 2304–2311.

Verron, S., Tiplica, T. & Kobi, A. (2007b). Multivariate control charts with a bayesian network, *4th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, pp. 228–233.

Verron, S., Tiplica, T. & Kobi, A. (2007c). Procedure based on mutual information and bayesian networks for fault diagnosis of industrial systems, *American Control Conference*, pp. 420–425.

Verron, S., Tiplica, T. & Kobi, A. (2008). Fault detection and identification with a new feature selection based on mutual information, *Journal of Process Control* **18**(5): 479–490.

Westerhuis, J., Gurden, S. & Smilde, A. (2000). Standardized q-statistic for improved sensitivity in the monitoring of residuals in mspc, *Journal of Chemometrics* **14**(4): 335–349.