Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Conditional Gaussian Network as PCA for fault detection ⋆

Mohamed Amine ATOUI * **Sylvain VERRON** *
Abdessamad KOBI *

* *LARIS/ISTIA, L'UNAM, 49000, Angers, France.*

Abstract: In this paper, we present a Bayesian framework for fault detection. Principal component analysis (PCA) technique and quadratic test statistics are incorporated under a Conditional Gaussian Network (CGN). The proposed network, given an observation, is able to project it into an orthogonal space and to give a decision about the system state (faulty or not). The paper demonstrate, the probability limits to use in order to match the decisions made by quadratic statistics. The equivalence between our method and PCA based fault detection is validated on the Tennessee Eastman process data sets.

Keywords: Fault detection, statistical inference, PCA, CGN, Tennessee Eastman process

1. INTRODUCTION

The growing demand for security and reliability of the current industrial processes (or systems), which become more and more complex, has made the Fault Detection and Isolation (FDI)-Fault detection and Diagnosis (FDD) an important and essential research topic. FDI-FDD comprises mainly two steps. First, the detection step seeks to identify at any moment if the system is in control (IC) or not. Once a change is confirmed (a fault has occurred: the system is Out of Control (OC)), the second step tries to explain it or to designate the responsible(s).

In the literature, we can find two main approaches for the FDD-FDI: model-based and data-driven methods. Theoretically, in the presence of an analytical representation of the system (detailed physical model), model-based methods are the best. However, obtaining this representation is often not possible, very tricky or request a lot of time and money. Unlike model-based methods, data-driven methods use only system measures (e.g. sensors, actuators) taken at different times (historical of data).

In this article, we focus only on data-driven methods for fault detection. A lot of techniques has been proposed in this research area Qin (2012); Yin et al. (2012); Venkatasubramanian et al. (2003). Many of them are based on rigorous statistical development of the process data, which generally employ multivariate/univariate test statistics (e.g. T^2, a special case of MEWMA (Multivariate Exponentially Weighted Moving Average), SPE (Squared Prediction Error), and so on). Among them we can mention Subspace Identification Methods (SIMs) and Subspace Aided Approach (SAP), tools developed to address the problems of building an accurate model for complex systems. They are directly designed utilizing the collected system data without explicitly identifying a system model (Ding, 2012). For fault detection purpose, techniques such as Fisher Disciminant Analysis, Independent Component Analysis (ICA), Partial Least Squares (PLS), Principal Component Analysis (PCA) and their variants could also be used. PCA is a well-known multivariate linear technique used in many fields due to its simplicity for model building and efficiency to handle a huge amount of process data (Tipping and Bishop, 1999).

Bayesian networks (BNs) have been also proposed for fault detection (Verron et al., 2010a; Kawahara et al., 2005; Schwall and Gerdes, 2002; Lerner et al., 2000). They are powerful tools dealing efficiently with multivariate models. BNs are designed by experts and/or learnt from data. We can cite some advantages to use them: Probabilistic/statistical frameworks, graphical representations of the dependencies between variables, the capacity to handle dynamics under dynamic BN's and possible fusion and integration of information from different sources.

In order to take better decisions, each information (e.g. probabilistic fault detection decision, maintainability information, components reliability and so on) on the system should be taken into account. In this perspective, we propose to use a BN in order to model PCA based fault detection. The major interests of this paper can be described in these few points: 1) a Gaussian representation of the PCA model 2) a generalization of the quadratic test statistics under a probabilistic tool, 3) for fault detection purpose, both PCA and quadratic statistics (T^2 and SPE) are managed under a single BN using discrete and Gaussian nodes (Bayesian framework).

The paper is structured in the following manner. In section 2, we introduce Bayesian Network (BN) and the particular one we use in this paper, we also briefly discuss PCA and its developments for fault detection. Section 3 presents an original CGN for fault detection. This is followed by a comparison study between the proposed network and PCA on the Tennessee Eastman process. Finally, conclusions and outlooks are given in the last section.

⋆ Mohamed Amine ATOUI is Supported by a PhD purpose grant from "la Région Pays de la Loire".

2. TOOLS

2.1 Bayesian Networks

A Bayesian Network (BN) is a probabilistic graphical model (Jensen and Nielsen, 2007), it is associated and consists of the following :

- a directed acyclic graph G, G=(V, E), where V and E are respectively its vertexes (nodes) and edges (arcs) sets,
- a finite probabilistic space $(\Omega, \mathbb{Z}, p)$, with Ω a non-empty space, $\mathbb{Z}$ a collection of the subspaces of Ω and, p a probability measure (we use the same notation for both probability distributions and probability density functions. The meaning will be clear from the context) on $\mathbb{Z}$ with $p(\Omega) = 1$,
- a set of random variables $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_l$ assigned to V and defined on $(\Omega, \mathbb{Z}, p)$, such that:
$$p(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_l) = \prod_{i=1}^{l} p(\mathbf{X}_i | p_a(\mathbf{X}_i) \quad (1)$$
where $p_a(\mathbf{X}_i)$ is the set of parent nodes of $\mathbf{X}_i$ in G,
- a conditional distribution associate to each node, given its parent nodes, describing probabilistic dependencies between variables,
- a calculations named inference, used given the availability of a new information (evidence) about one or several G nodes values, to update the network.

2.2 Conditional Gaussian Networks

One particular form of Bayesian networks is the Conditional Gaussian Network (CGN). Each node in the network represents a random variable that may be discrete or Gaussian (univariate/multivariate). However, to ensure availability of exact computation, discrete variables are not allowed to have continuous parents (see (Lauritzen and Jensen, 2001)). Thus, the Conditional Probability Distribution (CPD) for each discrete node given its parents, follows a multinomial distribution, outlined generally under a conditional probability table.

Unlike discrete nodes, Gaussian nodes are allowed to have Gaussian nodes as parents. Each Gaussian node, given its Gaussian parents follows a Gaussian linear regression model with parameters depending on the values of its discrete parents. In this paper, we restrict our attention to two kind of nodes.

First, the Gaussian linear node $\mathbf{Y}$, a Gaussian node with only parents $\mathbf{Z}_1, \ldots, \mathbf{Z}_c$, its conditional distribution is written as $p(\mathbf{Y}|\mathbf{Z}_1 = z_1, \ldots, \mathbf{Z}_c = z_c) = \mathcal{N}(\mu + W_1 z_1 + \cdots + W_c z_c; \Sigma)$, where μ is the parameter governing the mean and Σ the covariance matrix of $\mathbf{Y}$, $W_1, \ldots, W_c$ are the regression coefficients.

The second node is the Gaussian node $\mathbf{Y}$ having only discrete parents $\Pi = (\Theta_1, \ldots, \Theta_d)$. Its conditional distribution could be written as below for each value i_Π of its parents Π:

$$p(\mathbf{Y}|\Pi = i_\Pi) = \mathcal{N}(\mu_{i_\Pi}; \Sigma_{i_\Pi}), \; i_\Pi \in I_\Pi \quad (2)$$

where μ_{i_Π} and Σ_{i_Π} are respectively the mean of $\mathbf{Y}$ and its covariance matrix given the value i_Π of its parents. I_Π represent all the values that the parents of $\mathbf{Y}$ can take.

2.3 Principal Component Analysis

Principal component analysis (PCA) is a famous multivariate statistical technique (Jackson, 2005). Consider $X \in \mathbb{R}^{N\times m}$ a normalized set of N collected samples of input and output variables $\mathbf{x}$ of a given system. PCA maps linearly X, the original space, given a transformation matrix P to an orthogonal space T, and it separates it into two parts: the systematic part $\hat{X}$ and the noise part $\tilde{X}$.

$$T = XP, \; P \in \mathbb{R}^{m\times m}, \; P^T P = I \quad (3)$$

$$X = \hat{X} + \tilde{X} = \hat{T}\hat{P}^T + \tilde{T}\tilde{P}^T \quad (4)$$

The procedure to determine the matrix P and the two parts of X can be achieved as below:

Step I: Form the covariance matrix of X:

$$\Sigma \approx \frac{1}{N-1} X^T X \quad (5)$$

Step II: Find the eigenvalues λ_j and the eigenvectors P_j of Σ (see (Bishop et al., 2006)):

$$\Sigma = P\Lambda P^T, P = [P_1 \ldots P_m] \quad (6)$$
$$\Lambda = diag(\sigma_1^2, \ldots, \sigma_m^2), \; \sigma_1^2 \geq \cdots \geq \sigma_m^2 > 0, \; \sigma_j^2 = \lambda_j$$

Step III: Determine a (many methods exists see (Valle et al., 1999)), the number of dominant eigenvectors of Σ in which the retained variance under projection is maximal, and the principal axes $\hat{P} \subset P$ with the largest associated eigenvalues $\hat{\Lambda}$:

$$\Lambda = \begin{bmatrix} \hat{\Lambda} & 0 \\ 0 & \tilde{\Lambda} \end{bmatrix}, \hat{\Lambda} \in \mathbb{R}^{a\times a}, P = [\hat{P}, \tilde{P}], \hat{P} \in \mathbb{R}^{m\times a} \quad (7)$$

Step IV: Deduce $\hat{T}$ and $\tilde{T}$, and the parts of X:

$$T = [\hat{T} \; \tilde{T}], \; \hat{T} = X\hat{P}, \; \hat{T} \in \mathbb{R}^{N\times a}, \; \tilde{T} \in \mathbb{R}^{N\times(m-a)} \quad (8)$$

$$\hat{X} = \hat{T}\hat{P}^T, \; \tilde{X} = \tilde{T}\tilde{P}^T \quad (9)$$

Once P and a determined, PCA could be used for fault detection. It is generally associated to quadratic statistics (Ding et al., 2010). Based on T^2 and SPE statistic, fault detection using PCA could be done as below:

Step I: Define the control limits of T^2 and SPE for a given significance level α:

$$T^2 : CL_{T^2} = \frac{a(N^2-1)}{N(N-a)} F_\alpha(a, N-a) \quad (10)$$

$$SPE : CL_{SPE} = g\chi^2_{h,\alpha}, \theta_i = \sum_{j=a+1}^{m} (\sigma_j^2)^i, i = 1, 2 \quad (11)$$

$$g = \theta_2/\theta_1, h = \theta_1^2/\theta_2 \quad (12)$$

where χ^2 and F are respectively chi-squared distribution and Fisher distribution.

Step II: For each new and normalized (given the mean and variances of X) measurement sample $x \in \mathbb{R}^m$, compute T^2 and SPE as below:

$$T^2 = x^T \hat{P}\hat{\Lambda}^{-1}\hat{P}^T x = \hat{t}^T \hat{\Lambda}^{-1} \hat{t} \quad (13)$$

$$SPE = x^T (I - \hat{P}\hat{P}^T)^2 x = x^T \tilde{P}\tilde{P}^t x = \tilde{t}^T \tilde{t} \quad (14)$$

Step III: Finally, compare the calculated T^2 and SPE To their corresponding control limit and make a decision using this logic rule: if $T^2 \leq CL_{T^2}$ and $SPE \leq CL_{SPE}$ then the system is declared fault-free (IC), otherwise the system is faulty (OC).

2.4 Gaussian latent variable model and PCA

Based on a linear Gaussian model, (Tipping and Bishop, 1999) proposes a Probabilistic Principal Component Analysis (PPCA). It is a special case of statistical factor analysis and a generalisation of PCA (Kim and Lee, 2003). Indeed, consider X with row vector $x_n^T \in \mathbb{R}^m, n \in 1, \ldots, N$, using PPCA we can obtain, for example as below, the systematic part $\hat{X}$:

Step I: Consider and use the following probabilistic input-output model:

$$\mathbf{x} = \hat{A}\hat{\mathbf{t}} + \epsilon \tag{15}$$

$$\epsilon \sim \mathcal{N}(0, \Psi),\ \hat{\mathbf{t}} \sim \mathcal{N}(0, I), \mathbf{x}|t \sim \mathcal{N}(\hat{A}\hat{\mathbf{t}}, \Psi) \tag{16}$$

$$\Psi = \sigma^2 I,\ \sigma^2 \approx 0,\ \mathbf{x} \in \mathbb{R}^m,\ \hat{A} \in \mathbb{R}^{m \times a},\ \hat{\mathbf{t}} \in \mathbb{R}^a$$

where I is the identity matrix, ϵ is an $\mathbf{x}$-independent noise.

Step II: Using X, find a and the matrix $\hat{A}$ equivalent to $\hat{P} = [P_1, \ldots, P_a]$ (step II and III of PCA could be used, for other methods see (Bishop et al., 2006))

Step III: Calculate, given (15), the posterior distribution $p(\mathbf{t}|\mathbf{x} = x^n)$ according to:

$$p(\hat{\mathbf{t}}|\mathbf{x} = x_n) = \mathcal{N}(M\hat{A}^T x_n,\ \sigma^2 M),\ M \approx (\hat{A}^T \hat{A})^{-1} \tag{17}$$

$$\hat{t}^n = \mathbb{E}(\hat{\mathbf{t}}|\mathbf{x} = x_n) = M\hat{A}^T x_n \tag{18}$$

Step IV: Given $\hat{t}_n$, calculate $\hat{x}_n$:

$$\hat{x}_n = \mathbb{E}(\mathbf{x}|\hat{\mathbf{t}} = \hat{t}_n) = \hat{A}\hat{t}_n,\ \hat{x}_n^T \in \hat{X} \tag{19}$$

2.5 Discriminant Analysis and CGN

The discriminant analysis (DA) is a supervised statistic technique used to solve classification problem (see (Duda et al., 2001)), commonly under the assumption that the classes are normally distributed. Consider a new observation vector x of $\mathbf{x} \in \mathbb{R}^m$ and k different classes $C_{i, i \in 1, \ldots, k}$, DA affects x to the class C_i having the maximal a posteriori probability $p(C_i|x)$. This Maximum A Posteriori (MAP) rule, using the Bayes formula can be developed and written as below:

$$\begin{aligned}\delta : x \in C_{i^*}, \text{if } i^* &= \arg\max_{i=1,\ldots,k} p(C_i|x) \\ &= \arg\max_{i=1,\ldots,k} \frac{p(C_i)p(x|C_i)}{p(x)} \qquad (20) \\ &= \arg\max_{i=1,\ldots,k} p(C_i)p(x|C_i) \qquad (21)\end{aligned}$$

where $p(C_i)$ represents the a priori probability of the class $C_i, p(x)$ is the normalization factor which does not affect the decision and $p(x|C_i)$ is the multivariate normal probability density function of x given C_i:

$$p(\mathbf{x} = x|C_i) = \frac{1}{2\pi^{\frac{m}{2}}|\Sigma_i|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)} \tag{22}$$

where μ_i and Σ_i are respectively the mean vector and the covariance matrix of the class C_i, generally estimated using the Maximum Likelihood Estimation (MLE) given the available data.

From (20), many discrimination rules could be derived, often by making assumptions on classes covariance matrices. Quadratic Discriminant Analysis (QDA) arises when the covariance matrix of each class C_i is estimated given its corresponding training set. QDA could be represented in a CGN as shown in Figure 1. The network consists of a discrete root D, node that represent the k classes, and a multivariate Gaussian node $\mathbf{x} \in \mathbb{R}^m$ that takes into account correlation that may exist between the m variables. The D node value C_i with the maximum a posteriori probability is taken.

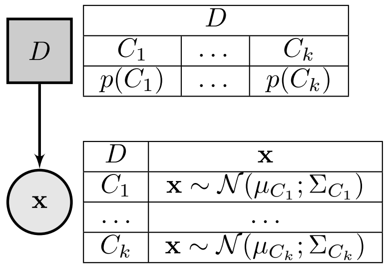

Fig. 1. CGN for DA

3. THE PROPOSED PROBABILISTIC FRAMEWORK

In this section, we present an original CGN for fault detection. The proposed network manages both PCA and quadratic statistics. First, we show how a linear projection (PCA) can be done under a CGN. After, we present the proposed probabilistic framework for test statistics as T^2 and SPE, and we give the final CGN for fault detection.

3.1 PCA and CGN

We have discussed in subsection 2.4 how PCA can be thought a generative model (see (15)), where all the variables are assumed Gaussian. This linear Gaussian model could be transposed under the CGN given in Figure 2. In this acyclic directed graph, the linear Gaussian node $\mathbf{x}$ correspond as in (15) to an m-dimensional observed variable that follows a conditional Gaussian distribution $\mathcal{N}(A\hat{\mathbf{t}}, \sigma^2 I)$, and the Gaussian parent node $\hat{\mathbf{t}}$ correspond to an a-dimensional Gaussian variable.

However, this CGN only manages the systematic part of PCA (see (9)) and the last $(m - a)$ components $\tilde{\mathbf{t}}$ of $\mathbf{t}$ are not handled. To make possible the separation of the original space into two subspaces, another Gaussian node $\tilde{\mathbf{t}}$, representing an $(m - a)$-dimensional variable is added. The structure of the new network is shown in Figure 3, it represents the following probabilistic model:

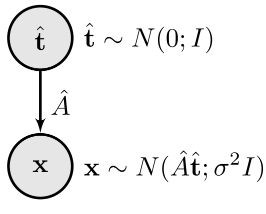

Fig. 2. PCA under a CGN

$$\mathbf{x} = \hat{A}\hat{\mathbf{t}} + \tilde{A}\tilde{\mathbf{t}} + \epsilon,\ \epsilon \sim N(0, \sigma^2 I) \tag{23}$$
$$\sigma^2 \approx 0, A = [\hat{A}\ \tilde{A}],\ \hat{A} = \hat{P},\ \tilde{A} = \tilde{P}$$
$$\mathbf{x} \sim N(\mu_{\mathbf{x}},\ \Sigma_{\mathbf{x}}),\ \mu_{\mathbf{x}} = \mathbb{E}(\epsilon) + \hat{A}\mathbb{E}(\hat{\mathbf{t}}) + \tilde{A}\mathbb{E}(\tilde{\mathbf{t}}) = 0 \tag{24}$$
$$\Sigma_{\mathbf{x}} = \hat{A}\mathbb{E}(\hat{\mathbf{t}}\hat{\mathbf{t}}^T)\hat{A}^T + \tilde{A}\mathbb{E}(\tilde{\mathbf{t}}\tilde{\mathbf{t}}^T)\tilde{A}^T + \mathbb{E}(\epsilon\epsilon^T) = \tag{25}$$
$$\hat{A}\hat{A}^T + \tilde{A}\tilde{A}^T + \sigma^2 I$$

3.2 T^2 and SPE under a CGN

In this subsection, we propose a probabilistic framework to handle multivariate/univariate quadratic statistics (we regroup them under the same notation Δ). Considering a new observation x of a given variable $\mathbf{x} \in \mathbb{R}^m$, a statistic Δ is first calculated and then compared to their predefined control limit CL_Δ. The system is declared out of control (OC), once it is upper to its CL_Δ. We want to find an equivalent rule using a CGN (see subsection 2.5), discriminating between the two classes: IC and OC.

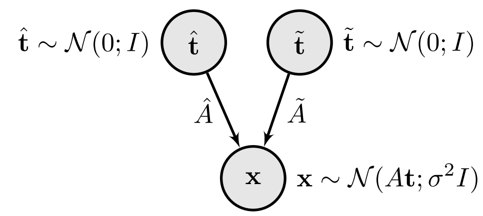

Fig. 3. PCA under a CGN

For that we have to find, for example, a probabilistic control limit (ζ_Δ^{oc}) such that: if the $p(OC|x) > \zeta_\Delta^{oc}$ the system is declared OC, where $p(OC|x)$ is the posterior probability of the class OC given a new observation x. To achieve this aim, the parameters (μ_{IC}, μ_{OC}, Σ_{IC}, Σ_{OC}) of the classes IC, OC need to be suitably defined.

The parameters of the class IC are estimated (if they are unknown) from the available free-fault data, using e.g. MLE. Regarding the class OC, as (Verron et al., 2010b) we consider it as a virtual class representing the set of observation that cannot be attribute to the IC class. Its parameters are defined so that $\mu_{\mathbf{x}} = \mu_{OC} = \mu_{IC}$ and Σ_{OC} express more variability than Σ_{IC} : $\Sigma_\Delta = \Sigma_{IC}; \Sigma_{OC} = \mathbf{c} \times \Sigma_\Delta$, where $\mathbf{c} > 1$ (if $\mathbf{c} = 1$, the two classes will be identical, which does not make sense). This difference ($\mathbf{c}$) making the variance of OC larger than IC is quite used in fault detection using Bayesian networks. It is generally estimated using faults data as in (Kawahara et al., 2005; Schwall and Gerdes, 2002; Lerner et al., 2000). However, faults data could not be available or not enough to accurately estimate it, which may lead increase false alarm and/or miss-detection rates. In this subsection, using only free-fault data, we propose a CGN able to provide decisions equivalent to those obtained by quadratic statistics. To achieve this, we shall seek probabilistic control limits (ζ_Δ^{oc}, ζ_Δ^{ic}) given any values of $\mathbf{c} > 1$ such as we keep the following decision rule:

$$x \in OC:\ if\ \Delta > CL_\Delta \tag{26}$$

under one of this decision rules:

$$x \in OC:\ if\ p(OC|x) > \zeta_\Delta^{oc} \tag{27}$$
$$x \in OC:\ if\ p(IC|x) \leq \zeta_\Delta^{ic} \tag{28}$$

Consider (28) and let:

$$p(IC|x) = \zeta_\Delta^{ic} \times 1 = \zeta_\Delta^{ic}[p(IC|x) + p(OC|x)] \tag{29}$$

where $\zeta_\Delta^{oc} = 1 - \zeta_\Delta^{ic}$. Using the Bayes formula, the a posteriori probability of each class (IC or OC) can be written as below:

$$p(D|x) = \frac{p(D)p(x|D)}{p(x)}, D \in \{IC, OC\} \tag{30}$$

from (29) and (30) we obtain:

$$\frac{p(IC)p(x|IC)}{p(x)} = \zeta_\Delta^{ic}\left[\frac{p(IC)p(x|IC) + p(OC)p(x|OC)}{p(x)}\right] \tag{31}$$

$$\zeta_\Delta^{ic} p(OC)p(x|OC) = p(IC)[p(x|IC) - \zeta_\Delta^{ic} p(x|IC)] \tag{32}$$

Let $\omega = \dfrac{p(OC)}{p(IC)}$, then we have:

$$\zeta_\Delta^{ic}\omega p(x|OC) = p(x|IC) - \zeta_\Delta^{ic} p(x|IC)$$
$$p(x|IC) = \zeta_\Delta^{ic}[p(x|IC) + \omega p(x|OC)]$$
$$\zeta_\Delta^{ic} = \frac{p(x|IC)}{p(x|IC) + \omega p(x|OC)} \tag{33}$$

As in QDA, each class, IC and OC, follows a Gaussian distribution, the conditional probability of this two classes can be written as in equation (34) and (35):

$$p(\mathbf{x}|IC) = \frac{\mathrm{e}^{\frac{-(\mathbf{x}-\mu_{\mathbf{x}})^T\Sigma_\Delta^{-1}(\mathbf{x}-\mu)}{2}}}{2\pi^{\frac{m}{2}}|\Sigma_\Delta|^{\frac{1}{2}}} \tag{34}$$

$$p(\mathbf{x}|OC) = \frac{\mathrm{e}^{\frac{-(\mathbf{x}-\mu_{\mathbf{x}})^T\Sigma_\Delta^{-1}(\mathbf{x}-\mu)}{2\mathbf{c}}}}{2\pi^{\frac{m}{2}}|\Sigma_\Delta|^{\frac{1}{2}}\mathbf{c}^{\frac{m}{2}}} \tag{35}$$

where p represents a multivariate Gaussian distribution of dimension m, Σ_Δ is the matrix assigned to $\mathbf{x}$ depending on Δ, and $(x - \mu_{\mathbf{x}})^T\Sigma_\Delta^{-1}(x - \mu_{\mathbf{x}})$ is the squared form of $\mathbf{x}$ when $p(IC|\mathbf{x} = x) = \zeta_\Delta^{ic}$. Hereinafter we refer to it by CL_Δ, then we have:

$$\zeta_\Delta^{ic} = \frac{\frac{\mathrm{e}^{(\frac{-CL_\Delta}{2})}}{2\pi^{\frac{m}{2}}|\Sigma_\Delta|^{\frac{1}{2}}}}{\frac{\mathrm{e}^{(\frac{-CL_\Delta}{2})}}{2\pi^{\frac{m}{2}}|\Sigma_\Delta|^{\frac{1}{2}}} + \omega\frac{\mathrm{e}^{(\frac{-CL_\Delta}{2\mathbf{c}})}}{2\pi^{\frac{m}{2}}|\Sigma_\Delta|^{\frac{1}{2}}\mathbf{c}^{\frac{m}{2}}}} \tag{36}$$

$$= \frac{1}{1 + \omega\frac{\mathrm{e}^{(\frac{\mathbf{c}-1}{2\mathbf{c}}CL_\Delta)}}{\mathbf{c}^{\frac{m}{2}}}} \tag{37}$$

Let $\gamma_\Delta = \dfrac{\omega}{\mathbf{c}^{\frac{m}{2}}}\mathrm{e}^{(\frac{\mathbf{c}-1}{2\mathbf{c}}CL_\Delta)}$ in (37), we finally obtain:

$$\zeta_\Delta^{ic} = \frac{1}{(1 + \gamma\Delta)},\ \zeta_\Delta^{oc} = 1 - \zeta_\Delta^{ic} \tag{38}$$

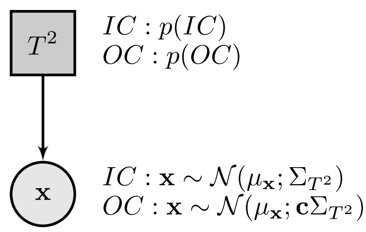

Fig. 4. T^2 under CGN

Once the probabilistic limits $\zeta_\Delta^{oc}, \zeta_\Delta^{ic}$ found, we are able given a $\mathbf{c}$ >1 to reproduce the test statistic Δ in a

CGN. In this paper, we require the probabilistic limits γ_Δ corresponding to T^2 and SPE. For that, it is sufficient to calculate $\gamma\Delta$, with the CL_Δ corresponding to each statistic.

Let $\Delta = T^2$, based on (10) the probabilistic control limits (38) can be obtained. Once $\zeta^{ic}_{T^2}$ or $\zeta^{ic}_{T^2}$ determined, we are able with the CGN presented in Figure 4 ,given a new observation x, to decide in which state the system belong: IC or OC.

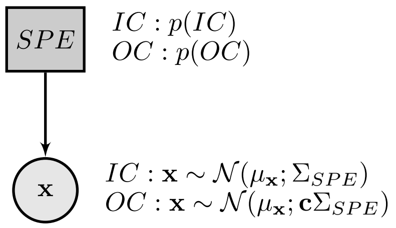

Fig. 5. SPE under CGN

SPE can be also done under a CGN similarly to T^2 (see Figure 5). Indeed, the Mahalanobis norm of $\mathbf{x}$ is identical to the Euclidean norm, when $\mathbf{x}$ is normalised $\Sigma_{SPE} = \Sigma_{T^2} = I$.

3.3 CGN for fault detection

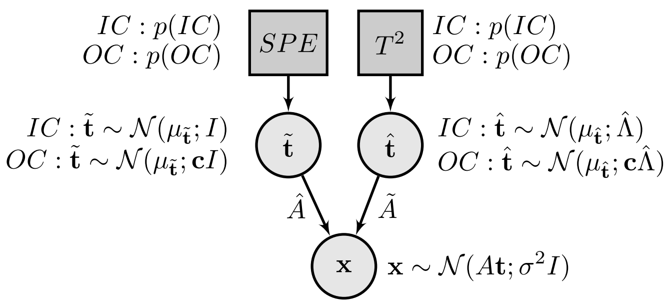

Fig. 6. PCA for fault detection under CGN

Previously, we have demonstrated the possibility to implement PCA and statistics test as T^2 and SPE under different CGN's. For fault detection purpose, we propose to join them under a single CGN (probabilistic framework) as shown in Figure 6. Given the structure of the proposed network, the nodes $\hat{\mathbf{t}}$ and $\tilde{\mathbf{t}}$ follows a conditional Gaussian distribution depending each on they discrete parents values IC, OC. Moreover, as they are respectively monitored by T^2 and SPE, their covariances matrices Σ_{T^2} and Σ_{SPE} are respectively defined equal to $\hat{\Lambda}$ (see (10)) and I (see (13)). Concerning the remaining nodes, the Gaussian node $\mathbf{x}$ follows a Gaussian linear regression model given its Gaussian latent parents $\hat{\mathbf{t}}$ and $\tilde{\mathbf{t}}$, and the discrete nodes (decisions nodes with two states IC, OC) are defined each given their states (values) prior probabilities.

4. APPLICATION

In this section, in order to compare our method to PCA (see section 2.3), we have tested it on the Tennessee Eastman Process (TEP). The TEP is an industrial chemical process, (see Figure 8) with five major units namely, reactor, condenser, compressor, separator and stripper as described in (Downs and Vogel, 1993). The process has 52 variables, 41 are the observed process variables and 11 are the manipulated variables. Its simulation provided by the Eastman Chemical Company is a well-known benchmark widely used for fault detection and/or diagnosis.

In this paper, as in (Yin et al., 2012) we consider 22 observed variables and 11 manipulated variables. Given the training fault-free data set (500 samples), 9 principal components are retained. The two methods are compared using the two indices: FAR (False Alarm Rate) and MDR (Miss Detection rate) on 21 test data sets (1 for normal operating conditions and others for 20 different faults). The network and its inferences was made under BNT (BayesNet toolbox (Murphy, 2001)). The results obtained shown that the both methods give same and identical results. In this paper, due to pages limitation, all the results can not be presented. However, Figure 7 shows the statistics T^2 and SPE and their equivalence under a CGN for the last 100 instants of the fault 4 data set. In this Figure, for the statistics T^2 and SPE, an upper violation of their respective control limit CL_{T^2} and CL_{SPE} means that a fault has occurred in the process. The opposite for the CGN, a lower violation of it respective probabilistic control limit ($\zeta^{ic}_{T^2}$and ζ^{ic}_{SPE}). It can be seen that the two approaches provide the same decisions at any instant.

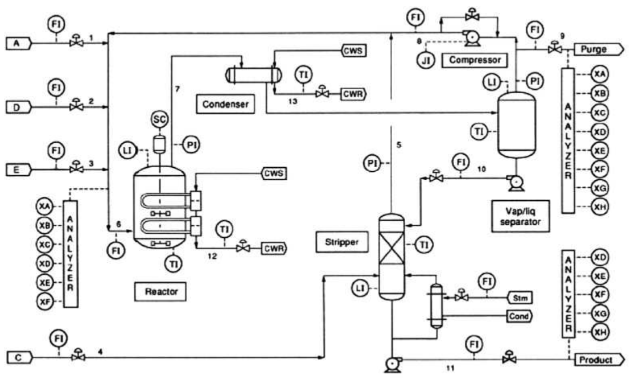

Fig. 8. Tennessee Eastman Processs

5. CONCLUSIONS AND OUTLOOKS

The main interest of this paper is the presentation of a new tool for fault detection. Firstly, we have transposed the PCA model in a BN, more precisely a CGN. Secondly, we have proposed an original probabilistic framework for squared statistics as T^2 and SPE. For that, it has been necessary to define probabilistic control limits in order to match the decisions made by the comparison of the squared statistics to their thresholds. Finally, we have integrated principal component analysis and its associated test statistics used for fault detection in a single CGN. The proposed method has been tested on the TEP and compared to PCA. The obtained results demonstrate that the two methods produce the same decisions.

This work lead to numerous outlooks such as handling missing values, integrating other information about the system e.g. data reliability, modeling others projection

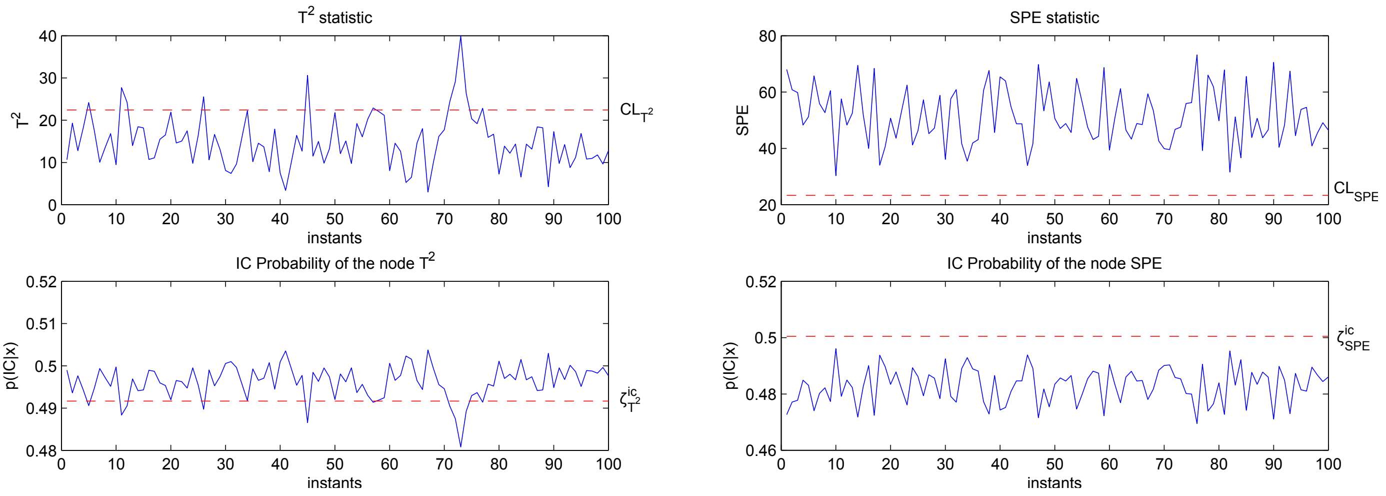

Fig. 7. Comparison between PCA and the proposed method (with $\mathbf{c} = 1.005$)

models and finally, managing temporal behaviors and non-Gaussian data hypothesis.

REFERENCES

Bishop, C.M. et al. (2006). *Pattern recognition and machine learning*, volume 1. springer New York.

Ding, S., Zhang, P., Ding, E., Yin, S., Naik, A., Deng, P., and Gui, W. (2010). On the application of pca technique to fault diagnosis. *Tsinghua Science & Technology*, 15(2), 138–144.

Ding, S.X. (2012). Data-driven design of model-based fault diagnosis systems. In *Proceedings of IFAC ADCHEM*, 10–13.

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

Duda, R.O., Hart, P.E., and Stork, D.G. (2001). *pattern classification 2nd edition*. Wiley.

Jackson, J.E. (2005). *A user's guide to principal components*, volume 587. Wiley. com.

Jensen, F.V. and Nielsen, T.D. (2007). *Bayesian networks and decision graphs*. Springer.

Kawahara, Y., Fujimaki, R., Yairi, T., and Machida, K. (2005). Diagnosis method for spacecraft using dynamic bayesian networks. In *'i-SAIRAS 2005'-The 8th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, volume 603, 85.

Kim, D. and Lee, I.B. (2003). Process monitoring based on probabilistic pca. *Chemometrics and intelligent laboratory systems*, 67(2), 109–123.

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

Lerner, U., Parr, R., Koller, D., and Biswas, G. (2000). Bayesian fault detection and diagnosis in dynamic systems. In *AAAI/IAAI*, 531–537.

Murphy, K.P. (2001). The bayes net toolbox for matlab. *Computing Science and Statistics*, 33, 2001.

Qin, S.J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2), 220–234.

Schwall, M.L. and Gerdes, J.C. (2002). A probabilistic approach to residual processing for vehicle fault detection. In *American Control Conference, 2002. Proceedings of the 2002*, volume 3, 2552–2557.

Tipping, M.E. and Bishop, C.M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), 611–622.

Valle, S., Li, W., and Qin, S.J. (1999). Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods. *Industrial & Engineering Chemistry Research*, 38(11), 4389–4401.

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

Verron, S., Li, J., and Tiplica, T. (2010a). Fault detection and isolation of faults in a multivariate process with bayesian network. *Journal of Process Control*, 20(8), 902 – 911.

Verron, S., Tiplica, T., and Kobi, A. (2010b). Fault diagnosis of industrial systems by conditional gaussian network including a distance rejection criterion. *Engineering Applications of Artificial Intelligence*, 23(7), 1229 – 1235.

Yin, S., Ding, S.X., Haghani, A., Hao, H., and Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of Process Control*, 22(9), 1567 – 1581.