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

Fault diagnosis based on graphical tools for multi-energy processes

R. Smaili * R. El Harabi * MN. Abdelkrim *

* Research Unit MACS, University of Gabès, National Engineering School of Gabès, Street Omar Ibn Elkhatab, 6029, Tunisie. (e-mail: rahma.smaili@gmail.com).

Abstract: To guarantee the safe operation of systems, it is necessary to use systematic techniques to detect and isolate faults for the purpose of diagnosis. The Fault Detection and Isolation (FDI) of nonlinear systems with coupling multiple energies became a difficult task.
This why, we propose in this paper, the exploitation of the behavioral and structural properties of graphs (Bond Graph and Signed Directed Graph) combining with a Principal Component Analysis method (PCA). Therein, a coupled Bond Graph model is used for modeling methodology. A Signed Directed Graph (SDG) is then deduced. Fault detection is later carried out by graph coloring. The localization of the actual fault is performed based on a nonlinear PCA (NLPCA) and back/forward propagations on the SDG.
The proposed approach is tested on the thermofluid case study and some simulations are provided.

Keywords: Coupled Bond Graph; Signed Directed Graph; Graph coloring; NLPCA; Fault Diagnosis; Thermo-fluid system.

1. INTRODUCTION

The fault detection and diagnosis are key issues for safe and optimal process operation. Existing literature is abundant with research works on fault diagnosis. Depending upon the knowledge used and the nature of information processing various techniques, for fault diagnosis, can be broadly categorized as a quantitative model-based method (Bond Graph (BG) ...), quantitative data-driven approach (principal component analysis, neural-networks...), qualitative model-based (signed directed graph, fault-tree ...). In this context, a SDG model captures both the information flow and the direction of effect (increase and decrease). Iri et al. (1979) were the first to introduce SDG for modeling chemical processes. Rule-based method using SDG has been used for fault diagnosis (FD) by kramer and palowitch (1987). Recently, Maurya et al. (2003) has proposed algorithms for the systematic development of SDG and digraphs for various types of systems and gave methodologies for SDG analysis to predict the initial and steady-state response of system which is very important for fault diagnosis. To improve the isolability of faults, this graph has been integrated with many approaches like the Principal Component Analysis (PCA) Vedam and Venkatasubramanian (1999), where a fault detection is performed using the PCA and the SDG model is involved to isolate the root causes. Yet, quantitative models, are described by mathematical relations and deduced from the fundamental laws (physical and chemical domains) and known as "residual methods". The BG tool, as a unified multi-energy domain, is especially suitable for developing models engineering processes and FDI. It was defined initially by Paynter (1961), this method allows covering of causal paths for the analysis of linear or nonlinear systems, modeled mono-energy Rahmani et al. (1997) or multi-energy systems Ould Bouamama (2005) with parameter uncertainties El Harabi (2011). Thus, a methodology to generate Analytical Redundancy Relations ($ARR's$) from a bond graph model is developed.
Methods based on principal component analysis could be very attractive for failure detection. This method can handle high dimensional and correlated process variables. In recent years, the PCA has been used in the statistical process control area such as process monitoring Raich and Cinar (1996) and sensor fault identification Dunia et al. (1996). However, principal component analysis is a linear approach, and most engineering problems are nonlinear. To overcome this problem, a nonlinear model to generate residuals for fault detection and isolation is given. Hastie and Stuetzle (1989) proposed a principal curve methodology to provide a nonlinear summary of a m-dimensional data set. However, this approach is non-parametric and can not be used for continuous mapping of new data. To overcome this parametrization problem, an auto-associative neural network Kramer (1991) has been used.
After all in our previous works Smaili et al. (2012b), Smaili et al. (2012a) diagnosis based on BG is done through generation $ARR's$ which are a complex task, fault detection and isolation based on SDG method is made by getting initial or steady-state response and propagation through paths which can generate a spurious solutions that can not indicate the actual fault. To overcome these problems, we propose a new fault detection and isolation algorithm. The main contribution of the present work is the development of a combined nonlinear dynamic PCA and graphical approaches-based fault diagnosis algorithm. The proposed approach concerns use of coupled bond graph models for

modeling using the behavioral, structural and causal properties of an integrated graphical tool. A signed directed graph is hence deduced. Fault detection is later performed using initial responses of all measured variables. Whenever an abnormality is indicated, a nonlinear dynamic PCA and back/forward propagations through paths from exogenous variables to system variables on SDG model are combined so as to identify fault roots.
The rest of this paper is organized as follows: Section 2 presents briefly a graphical representation. After that, in Section 3, nonlinear PCA and its interest for FDI is given. The integrated design scheme combining graphical and nonlinear PCA approaches of FD system is given later. Then, the efficiency of the proposed methods for monitoring is shown through a case study described in section 5. The concluding remarks and future scope of works are mentioned in section 6.

2. GRAPHICAL REPRESENTATION

Before starting the description of the proposed algorithm, it is necessary to introduces same notations used in this paper.

2.1 Graph

In graph theory, the graph G is defined as a dual pair (S, A) where $S = \{s_1...s_i...s_n\}$ is the node set of graph G, $A = \{a_1...a_i...a_n\}$ is the edge set of graph. Therefore, a function $f : (S, A) \to \Xi$ where $\Xi \in \mathbb{R}^+$ gives the wight of S and A.

2.2 Graph coloring

The graph coloring is the way of coloring the nodes of a graph such that no two adjacent nodes share the same color. For a given graph G, a function $c_s : S \to 1...k$ such that $\forall(s_i, s_j) \in A$, $c_s(s_i) \neq c_s(s_j)$ is called the color of s_i. Similarly, an edge coloring assigns a color to each edge so that no two adjacent edges share the same color.
otherwise, color of nodes can be classified into three colors:

- white: when s_i is unvisited;
- gray: when s_i is visited and $c_s(succ(s_i))$ = white;
- black: when s_i is visited and $c_s(succ(s_i))$ = gray or black.

2.3 Breadth First Search and Depth First Search

- The Breadth First Search (BFS) begins at a root node and inspects all the neighboring nodes. Then for each of those neighbor nodes in turn, it inspects their neighbor nodes which were unvisited, and so on. From the standpoint of the algorithm, all child nodes obtained by expanding a node are added to a FIFO (First In, First Out) queue.
- The Depth First Search (DFS) is an algorithm for traversing or searching a tree, tree structure, or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

2.4 Signed Directed Graph

The signed directed graph is a particular case of graphs where functions $f_s : S \to \Xi$, $\Xi \in \{+, 0, -\}$ and $f_A : A \to \Xi$, $\Xi \in \{+, -\}$ are the sign of nodes and edges respectively. Nodes in the SDG represent system variables: input, output and state variables, its sign represent the qualitative state, indeed, "0" is the normal state, "+" and "-" indicate, respectively, if this variable is increasing or decreasing from it normal state. Edges of the SDG modeled cause and effect relations between different variables. "+" and "-" means that cause and effect change respectively in the same sense or in the opposite sense.

2.5 Bond Graph

The Bond Graph is a graph where its nodes represent BG elements and edges show the power transfers within a system. BG is unified for all physical fields Dauphin-Tanguy (2000). A detailed presentation of the bond graph method is contained in Dauphin-Tanguy (2000) Ould Bouamama (2005).
In process engineering processes, several phenomena (chemical, thermal and fluidic) are coupled. In addition to matter transformation phenomena, chemical and electrochemical processes involve additional complexity in the modelling task, since the mass that flows through the process carries the internal energy which is stored in it, and which is thus transported from one location to another in a non dissipative fashion. Power variables are thus in vectorial form Ould Bouamama (2002):

$$E = [e_h \;\; e_t \;\; e_c]^T \tag{1}$$

$$F = [f_h \;\; f_t \;\; f_c]^T \tag{2}$$

where e_h, e_t and e_c represent respectively the thermal effort (specific enthalpy h or the temperature T), the hydraulic effort (the pressure P), and the chemical effort (the chemical potential μ, chemical affinity A or the concentration c). f_h, f_t and f_c represent respectively the thermal (or entropy) flow, hydraulic flow and chemical flow (molar flow $\dot{n}$).

3. PCA-BASED PROCESS DIAGNOSIS

The principal component is defined as a linear transformation of the original variables into a new set of variables which are uncorrelated to each other. PCA involves the decomposition of a data matrix $X \in \mathbb{R}^{m \times n}$, which contains m sampled observations of n process variables, into a transformed subspace of reduced dimension. X is assumed to be normally distributed with zero mean and unit variance. The covariance matrix of X is defined as:

$$\Sigma = \frac{1}{m-1} X^T X \tag{3}$$

The linear transformation is:

$$T = XP \tag{4}$$

where $T \in \mathbb{R}^{m \times n}$ is the principal component matrix, and the matrix $P \in \mathbb{R}^{n \times n}$ contains the principal vectors which are the eigenvectors associated with the eigenvalues λ_i (in decreasing magnitude order) of the covariance matrix Σ.
The partition of eigenvectors and principal component matrices are given as follows:

$$P = \begin{bmatrix} \hat{P} & \tilde{P} \end{bmatrix}, T = \begin{bmatrix} \hat{T} & \tilde{T} \end{bmatrix} \tag{5}$$

Therefore, equation 4 becomes:

$$X = \hat{T}\hat{P}^T + \tilde{T}\tilde{P}^T = \hat{X} + E \tag{6}$$

The matrices $\hat{X} = X\hat{C}$ and $E = X\tilde{C}$ denote, respectively, the modelled variations and nonmodelled variations of X based on l components ($l < m$).
with the matrices $\hat{C} = \hat{P}\hat{P}^T$ and $\tilde{C} = I - \hat{C}$ constitute the PCA model.
Fault detection using PCA is performed by monitoring the residuals. There are two detection index: the SPE (squared prediction error) is a statistic that measures the lack of fit of the PCA model to the data and the Hotelling's T^2 represents the major variation of the data. At time k, the detection index SPE and T^2 are given, respectively, by:

$$SPE(k) = \sum_{i=1}^{m} (e_i(k))^2 \tag{7}$$

$$T^2(k) = \sum_{i=1}^{l} \frac{t_i^2(k)}{\lambda_i} \tag{8}$$

These quantities suggest the existence of an abnormal situation in the data when:

$$SPE(k) > \delta_\alpha^2 = \theta_1 \left(1 + \frac{c_\alpha h_0 \sqrt{2\theta_2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2}\right)^{\frac{1}{h_0}} \tag{9}$$

$$T^2(k) > T_\alpha^2 = \chi_{l,\alpha}^2 \tag{10}$$

where δ_α^2 and T_α^2 are, respectively, control limit for SPE and T^2. $\chi_{l,\alpha}^2$ is a distribution with l degree of freedom. $\theta_i = \sum_{j=l+1}^{n} \lambda_j^i$ and $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$.
When the fault is occurred and detected, it is important to identify fault roots and apply the necessary corrective actions to eliminate the abnormal data. Hence, contribution plots Miller et al. (1998) are used to identify the faulty variables.
The contribution of process variable j to the SPE-statistic at time period k is:

$$cont_j^{SPE}(k) = (e_j(k))^2 = (x_j(k) - \hat{x}_j(k))^2 \tag{11}$$

In the case of T^2-statistic, the contribution of the process variable x_j for a normalized principal component $\left(\frac{t_i}{\sigma_i}\right)^2$ (σ_i is a singular value equal to $\sqrt{\lambda_i}$) is:

$$cont_{i,j} = \frac{t_i}{\lambda_i} p_{i,j} x_j \tag{12}$$

Thus, the total contribution to the T^2-statistic of a variable x_j is as follows:

$$Cont_j = \sum_{i=1}^{l} cont_{i,j} \tag{13}$$

Consequently, a process variable is identified as a fault when it has the higher contribution plot.

3.1 Dynamic PCA method

Dynamic principal component analysis (DPCA) proposed by Ku Ku et al. (1995) aims at finding dynamical linear relations between the process variables which consider a t_l time-lagged, this is done by combining the observations of the measurement vector with the s previous observations in the data matrix as follows:

$$X(t_l) = [X(k) \;\; X(k-1) \;\; \cdots \;\; X(k-s)]$$
$$= \begin{bmatrix} x^T(k) & x^T(k-1) & \cdots & x^T(k-s) \\ x^T(k-1) & x^T(k-2) & \cdots & x^T(k-s-1) \\ \vdots & \vdots & \ddots & \vdots \\ x^T(k+s-n) & x^T(k+s-n-1) & \cdots & x^T(k-n) \end{bmatrix} \tag{14}$$

The procedure for selecting t_l is discussed in detail in Ku et al. (1995).

4. PROPOSED ALGORITHM

The new algorithm presented in this section is dedicated to fault detection and isolation for nonlinear systems, in which the BG and SDG and NLPCA tools are combined. Overall flow of the proposed algorithm (Fig 1) is explained below:

(1) Modelling step: it is to get graphical model systems by developing a signed directed graph model directly from the bond graph (Table 1).
(2) Fault detection: by using the graph coloring and the Breadth First Search in the SDG model that deduced, a fault is detected when the non-zero sign of a measured variable, due to changes in an exogenous variable, is given.
The Breadth First Search is complete when all measured variables are visited and their sign are determined.
(3) Fault isolation: when a fault is detected, contribution plots of process variables are determined using a NLPCA, the variable with a higher contribution is the faulty variable. So, it is selected and if it is a measured node in the SDG, then the backward-propagation from this variable to the fault node is performed. If it was an exogenous variable then it constitutes a candidate fault.

5. CASE STUDY

The case study deals with a thermo-fluid system Noura (2009) depicted in Fig. 2, where Q_0 and T_0 are respectively the output flow and the temperature of the outlet fluid.
Parameters values are defined in Table 2.
The coupled bond graph model in integral causality is depicted in Fig. 3. The two ports C_{ht} represent the coupled thermal and hydraulic energies of the stored fluid (considered here in under saturated state) is decoupled into thermal and hydraulic capacity C_t and C_h. $S_f : P_u$, $S_f : Q_i$ and $S_e : T_i$ modelled, respectively, a thermal flow source, an inlet mass flow, and a temperature of the inlet fluid (considered constant). The coupling is modelled by the fictive R_{c1} and R_{c2} elements in thermal bonds.
Based on BG-SDG analog (see Table 1), the signed directed graph model of the thermo-fluid system is deduced directly from a coupled bond graph model, as shown in Fig. 4.
Now, we consider 9 faults (actuator, sensor and process) affected the system. These faults are listed in Table 3.
Initially all edges are white and all the nodes are in the normal state ($f_s(S_i) = 0$). Nodes of degree $d^- = 0$ are Q_i T_i, R. If we start by a tank leakage $R(R+)$ then

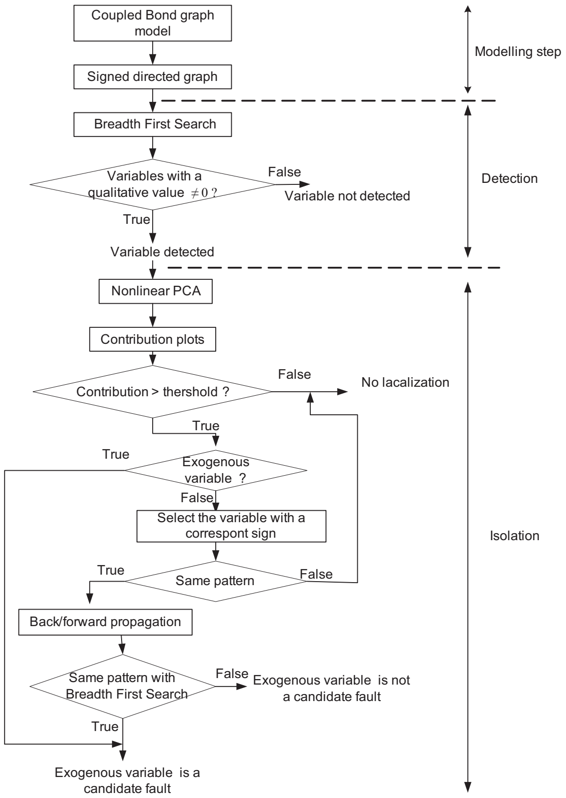


Fig. 1. Proposed algorithm.

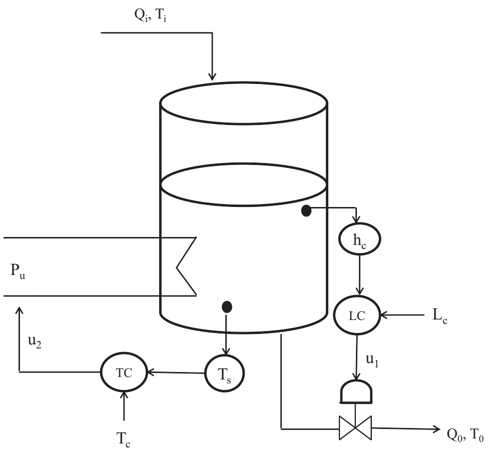


Fig. 2. Thermo-fluid system

$succ(R) = h$, $C_A(R, h) = gray$. So, $h = -$: as arcs between R and h is negative. Then $succ(h) = h_c$ and T_0 and Q_0. $C_A(h, hc) = gray$ and $C_A(h, Q_0) = gray$ and $C_A(h, T_0) = gray$. Indeed, $h_c = -$ and $Q_0 = -$ and $T_0 = +$. At this level signs of system variables $h\ T_0\ Q_0$ are determined, see Fig. 5, then the algorithm is complete and therefore was obtained symptom of these variables following a decrease of Q_i, this symptom is $[h\ \ T_0\ \ Q_0] = [-\ +\ -]$.

To isolate faults, the NLPCA is used. The case study has six measured variables (Q_i, P_u, T_i, h, T_0 and Q_0), therefore, the vector $x(k)$ is $x(k) = [Q_i\ \ P_u\ \ T_i\ \ h\ \ T_0\ \ Q_0]^T$. Here, only $m = 61$ samples are used. The proposed approach is based on a neuronal NLPCA model with five layers with three hidden layers. We have chosen to extract

Table 1. BG-SDG analog

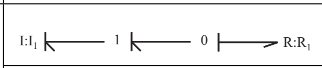
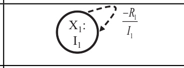
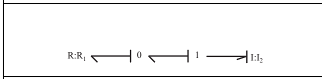
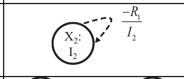
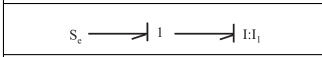
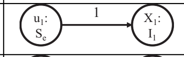
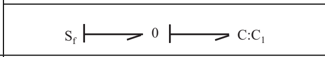
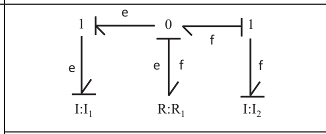
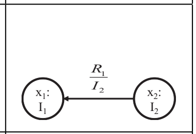
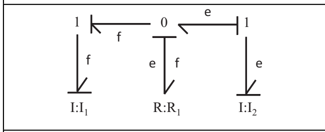
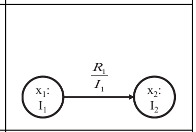
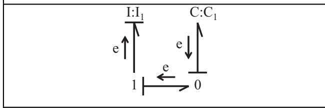
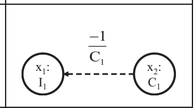
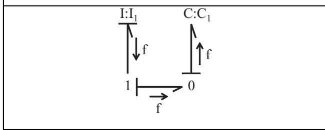
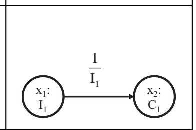
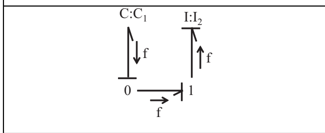
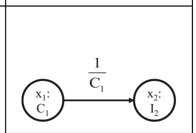
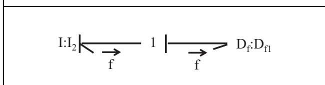
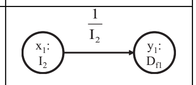
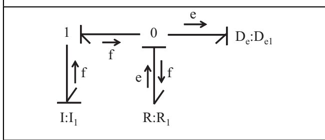
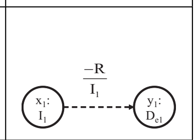
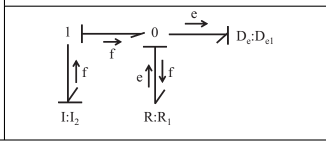
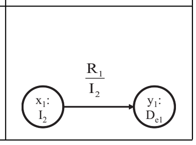
Path	BG element	Gain	SDG element
L_0		$\frac{-R_1}{I_1 s}$	
		$\frac{-R_1}{I_2 s}$	
		1	
		1	
L_1		$\frac{R_1}{I_2 s}$	
		$\frac{R_1}{I_1 s}$	
		$\frac{-1}{C_1 s}$	
		$\frac{1}{I_1 s}$	
		$\frac{1}{C_1 s}$	
		$\frac{1}{I_2 s}$	
		$\frac{-R_1}{I_1 s}$	
		$\frac{R_1}{I_2 s}$	

Table 2. parameter values

Parameter	Symbol	Value
Flow rate at the inlet	Q_i	0.7 m^3/s
Cross-section area of tank	S	1 m^2
Heat capacity	c	4.2 $J.g^{-1}.K^{-1}$
Density of water	ρ	1000 $g.l^{-1}$
Constant which depend of the valve	α	0.3
Electric power	P_u	2 KW

the nonlinear principal components by parallel way. In this example, three nonlinear components ($l = 3$) have been retained with $t_l = 1$.

When a leakage in the tank ($R(+)$) is produced, the contribution of process variables to T^2-statistic, as shown in Fig. 6, indicate that only the variable 4 (h) has the highest contribution with a negative sign. This variable is selected and as h is a measured variables in the SDG model, therefore, the backward propagation from this variable is performed. By exploiting causal paths between

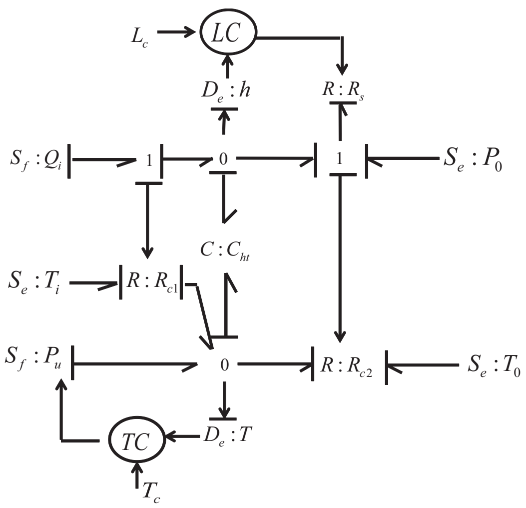


Fig. 3. Coupled bond graph model of a thermo-fluid system

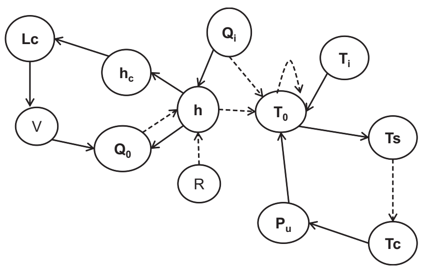


Fig. 4. SDG of a thermo-fluid system

Table 3. Fault list

Fault n	Description	Type of faults	Symbol
1	Q_i high	Actuator	Q_i(+)
2	Q_i low	Actuator	Q_i(-)
3	P_u high	Actuator	P_u(+)
4	P_u low	Actuator	P_u(-)
5	T_i high	Actuator	T_i(+)
6	T_i low	Actuator	T_i(-)
7	h_c high	Sensor	h_c(+)
8	T_c high	Sensor	T_c(+)
9	Leak in the tank	Process	R(+)

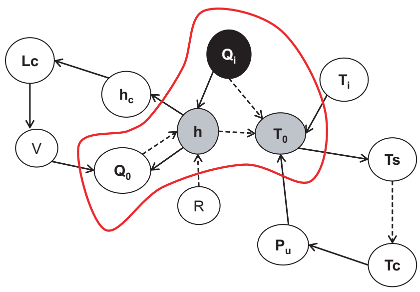


Fig. 5. SDG of a thermo-fluid system with nodes coloring

variables of the SDG, the predecessor variables of h are R and Q_i. By affecting these exogenous variables with signs, we can get the leakage in the tank $R(+)$ and a low variation of the flow rate at the inlet $Q_i(-)$. The forward propagation from these two variables gives the same pattern of measured variables found previously. Consequently, $R(+)$ and $Q_i(-)$ are isolated as a possible fault, but $R(+)$ is the candidate fault.

The same result is obtained by the contribution to SPE-

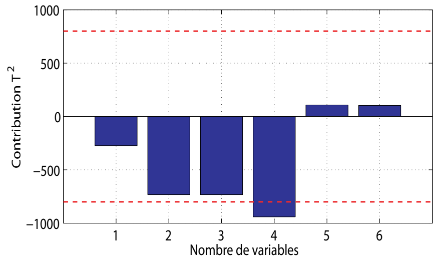


Fig. 6. Contribution of process variables to T^2-statistic.

statistic (Fig. 7), the variable with a higher contribution (h) is selected.

We conclude that fault diagnosis of each process is

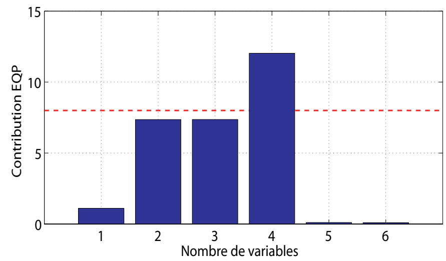


Fig. 7. Contribution of process variables to SPE-statistic.

performed using the proposed method. The result of a fault isolation is consistent when actuator fault is occurred. But whenever a process fault (leakage in the tank $R(+)$) two possible faults are determined, hence this drawback of the present algorithm is due to the complexity of each system and as known of the SDG approach is a qualitative one that need more information about faults.

6. CONCLUSION

An automated framework for the interpretation of causal graphical approaches (coupled BG and SDG) using a non-linear dynamic PCA to perform process monitoring and diagnosis has been developed.
An advantage of this approach is the automation of SDG-based fault diagnosis where the situation ambiguities to determine the faulty variables are replaced by automated interpretation of the contribution plots using nonlinear dynamic PCA. The proposed algorithm also reduced the number of spurious solutions of the SDG-based fault diagnosis, in which a faulty variable is selected by nonlinear PCA and the use of back/forward propagation on the SDG. Efficiency of this combination is illustrated on the thermo-fluid system which make in evidence the interaction of two energies (hydraulic and thermal), so that a coupled bond graph is suitable tool to model this kind of processes.
In the future work algorithms for the diagnosis of uncertain systems based on the BG-LFT model will be developed.

REFERENCES

Dauphin-Tanguy, G. (2000). *Les bond graphs*. Hermès science publications.

Dunia, R., Qin, S., Edgar, T., and McAvoy, T. (1996). Identification of faulty sensors using principal component analysis. *AIChE Journal*, 42(10), 2797–2812.

El Harabi, R. (2011). *Supervision des Processus Chimiques Base de Modles Bond Graph*. Ph.D. thesis.

Hastie, T. and Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84(406), 502–516.

Iri, M., Aoki, K., O'Shima, E., and Matsuyama, H. (1979). An algorithm for diagnosis of system failures in the chemical process. *Computers & Chemical Engineering*, 3(1-4), 489–493.

Kramer, M. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal*, 37(2), 233–243.

kramer, M. and palowitch, B. (1987). A rule-based approach to fault diagnosis using the signed directed graph. *AIChE Journal*, 33(7), 1067–1078.

Ku, W., Storer, R., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1), 179–196.

Maurya, M., Rengaswamy, R., and Venkatasubramanian, V. (2003). A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. algorithms and analysis. *Industrial & engineering chemistry research*, 42(20), 4789–4810.

Miller, P., Swanson, R., and Heckler, C.E. (1998). Contribution plots: a missing link in multivariate quality control. *Applied mathematics and computer science*, 8, 775–792.

Noura, H. (2009). *Fault-tolerant control systems: Design and practical applications*. Springer.

Ould Bouamama, B. (2002). *Modélisation et supervision des systèmes en Génie des procédés–approche Bond Graphs*. Ph.D. thesis, Laboratoire d'Automatique et Informatique Industrielle de Lille USTL, France.

Ould Bouamama, B. (2005). Applications de la méthode bond graph à la modélisation des systèmes énergétiques. *Article publié dans les techniques d'ingénieurs*.

Paynter, H. (1961). *Analysis and design of engineering systems*. MIT press.

Rahmani, A., Sueur, C., and Dauphin-Tanguy, G. (1997). Approche des bond graphs pour l'analyse structurelle des systèmes linéaires. *Linear Algebra and its applications*, 259, 101–131.

Raich, A. and Cinar, A. (1996). Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. *AIChE Journal*, 42(4), 995–1009.

Smaili, R., EL Harabi, R., and Abdelkrim, M. (2012a). Fdi based on causal graphical approaches for nonlinear processes. In *10th International Multi-Conference on Systems, Signals and Devices SSD'2013, Hammamet (TUNISIA)*. IEEE.

Smaili, R., EL Harabi, R., and Abdelkrim, M. (2012b). Model-based process diagnosis: Bond graph and signed directed graph tools. In *International Conference on Control, Decision and Information Technologies CoDIT'13, Hammamet (TUNISIA)*. IEEE.

Vedam, H. and Venkatasubramanian, V. (1999). Pca-sdg based process monitoring and fault diagnosis. *Control Engineering Practice*, 7(7), 903–917.