DYNAMICAL CLUSTERING TECHNIQUE TO ESTIMATE THE PROBABILITY OF THE FAILURE OCCURRENCE OF PROCESS SUBJECTED TO SLOW DEGRADATION

M. Traore, E. Duviella and S. Lecoeuche
Departement Informatique et Automatique, Ecole des Mines de Douai, France

Keywords: Supervision, Pattern Recognition, Non-stationary data, AUDyC.

Abstract: In this paper, we propose a supervision method which aims at determining pertinent indicators to optimize predictive maintenance strategies. The supervision method, based on the AUto-adaptative and Dynamical Clustering technique (AUDyC), consists in classifying in real time measured data into classes representative of the operating modes of the process. This technique also allows the detection and the tracking of the slow evolutions of the process modes. Based on the AUDyC technique, a method is proposed to estimate the probabilities of the failure occurence of components in real time. This method is illustrated on the real case of a temperature controller.

1 INTRODUCTION

Maintenance strategies consist in improving the safety and the reliability of industrial processes, taking into account their characteristics and the cost of maintenance plans (Grall et al., 2002). Amongst the three principal types of maintenance strategies which are proposed in the literature (Muller et al., 2004), *i.e.* the corrective, the preventive and the predictive maintenance strategies, the predictive maintenance allows the anticipation of failures and the optimal selection of maintenance actions, by the estimation in real time of the current state of the process components. This strategy is generally based on supervision methods and the estimation of the failure occurrence probabilities of the components of the process. The initial selection of the components which are essential to supervise, is performed by a dysfunctional analysis of the failure modes and their effects (FMEA: Failure Mode and Effects Analysis). Then, the interactions between each component are modelled by a Fault Tree formalism (Lassagne, 2000), (Vesely et al., 1981). Finally, the Fault Tree can be quantified by using the concept of Probability Functions by Episode (PFE) which allow the association of a probability of occurrence function to each component. In (Desinde et al., 2006), the PFE of the components are supposed to be known *a priori* and resulted from factory tests of feedback methods. We propose in this paper a supervision method allowing of determine the PFE in real time. The supervision methods based on mathematical models of the process can not be used for complex processes or when no physical model is available. In these cases, supervision approaches which consist in extracting relevant and sensitive informations of the component state by using directly the sensor signals, are more efficient. These supervision methods gather Pattern Recognition (PR) techniques which involve the state of a component by the analysis of evolutive data. The PR techniques include for exemple dynamic classification algorithms for evolutive data defined in (Lurette and Lecoeuche, 2003), which are dedicated to associate a state to one of the several operating modes of the system. FMMC (Min-Max Fuzzy Clustering) (Mouchawed and Billaudel, 2002) or AUDyC (AUto-adaptive and Dynamical Clustering) techniques allow the detection and the tracking of fast and slow evolutions of non-stationary data, and the diagnosis of the current state of the process. AUDyC approach is specially adapted to the supervision of slow evolutions or drifts due for exemple to ageing phenomenon (Lecoeuche et al., 2004). It allows the classification of the observed data according to classes which correspond to the operating modes of the process, *i.e.* normal, current and default modes. Estimation techniques of the distances between the several classes have been proposed to quantify the positioning of each classe. In this context, the main difficulty is to estimate the probabilities of the failure occurrence of components according to the dy-

Traore M., Duviella E. and Lecoeuche S.
DYNAMICAL CLUSTERING TECHNIQUE TO ESTIMATE THE PROBABILITY OF THE FAILURE OCCURRENCE OF PROCESS SUBJECTED TO SLOW DEGRADATION.
DOI: 10.5220/0002250003600365
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2009)*, page
ISBN: 978-989-674-000-9

namic data classification, and finally to provide indicators allows the improvement of predictive maintenance strategies.

In this paper, we consider processes characterised by slow evolutions of their operating modes. We propose to use AUDyC technique to supervise the components of the process, to estimate the probability of occurrence of each component of the process. The problematic addressed in this paper is detailled in the section 2. The supervision method by AUDyC is presented in Section 3. In Section 4, we present the methods proposed to estimate the probability of the failure occurrence of components of the process. Finally, the proposed methods are applied to a temperature controller.

2 PROBLEMATIC

Considering processes subjected to slow drifts of their current mode towards default modes, we propose a method which aims at determining indicators like probabilities of the failure occurrence of components. These indicators can be used to optimize predictive maintenance plans. The first step of maintenance strategy consists in a FMEA of the process to determine the corresponding Fault Tree, to specify the elementary component and the interactions between each component. The FMEA of the process leads also to the determination of components which are necessary to be supervised. The Fault Tree is quantified by using Probability Functions by Episode (PFE) (*see* Figure 1), where $PFE(E_x)$ which denotes the PFE of the event E_x is expressed by relation (1). The PFE of events associated to elementary components, *i.e.* E_1 to E_4, are used to compute the PFE of others events, E_5 and E_6.

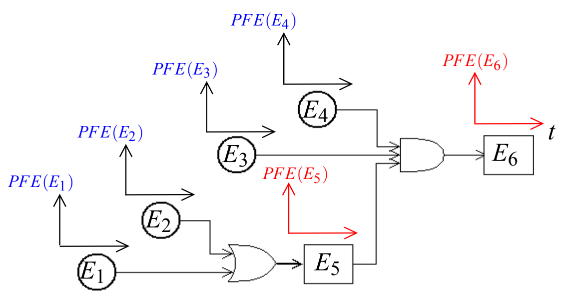

Figure 1: Fault Tree and PFE associated to components.

$$PFE(E_j) = ((p_1^{E_j}, t_1), \cdots, (p_n^{E_j}, t_n)) \qquad (1)$$

$\forall\, t_i \quad p_i^{E_j} = p^{E_j}(t_i)$, where $p^{E_j}(t_i)$ is the failure occurrence probability of the event E_j of the component j at time t_i.

The components which have to be monotored being known, it is necessary to select the variables which are characteristics of the component state. Three states are considered: normal, current and default modes. The goals of the dynamic data classification technique is to classify the measured data according to normal, current or default classes in real time. The estimation of characteristics of the current class leads to the detection and the tracking of drifts. The normal and default classes are known *a priori*, and are represented in the data representation space (*see* Figure 2). The slow drift of an operating mode has for effect of gradual change of the data from the normal class to the default class. The goal is to characterize in term of PFE the drift of an operating mode from the normal mode to the default mode. For that, we use AUDyC technique as modelling technique and estimation techniques of the distances between classes, as Euclidean and Kullback-Leibler distances. The AUDyC technique and the estimation methods of the distances are presented in the next section.

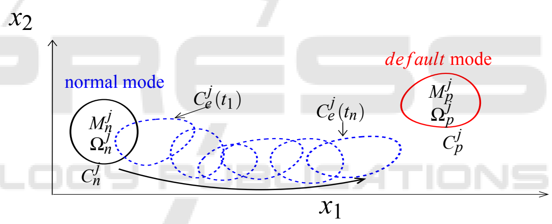

Figure 2: Slow drift operating.

3 CURRENT CLASS MODELLING BY AUDYC TECHNIQUE

The supervision method based on the AUDyC technique aims at monitoring each component of the process and at determining their mode. An operating mode is represented by a Gaussian class C_k^j which is characterized by a center M_k^j and a matrix of covariance Ω_k^j. These parameters are estimated in real time according to the observed data contained into the observation vector which is denoted $X^i = [x_1^i, x_2^i, \cdots, x_d^i]$ in the d space dimensions. The AUDyC algorithm consists in updating the class parameters recursively on a sliding window of width N_{fen} taking into account the cardinality of the class C_k^j, *i.e.* $Card(C_k^j)$. The steps of the algorithm, detailed in (Lecoeuche et al., 2004), are presented thereafter:

- If $Card(C_k^j)$=nb ¡ N_{fen}: Add information

$$M_k^j(t) = M_k^j(t-1) + \frac{1}{nb+1}(X(t) - M_k^j(t-1))$$
$$\Omega_k^j(t) = \frac{nb-1}{nb}\Omega_k^j(t-1) +$$
$$\frac{1}{nb+1}(X(t) - M_k^j(t-1))^\top (X(t) - M_k^j(t-1)) \tag{2}$$

- If nb $\geq N_{fen}$: Add and remove information

$$M_k^j(t) = M_k^j(t-1) + \frac{1}{N_{fen}}(\delta X^+ - \delta X^-)$$
$$\Omega_k^j(t) = \Omega_k^j(t-1) +$$
$$\Delta X \begin{bmatrix} \frac{1}{N_{fen}} & \frac{1}{N_{fen}(N_{fen}-1)} \\ \frac{1}{N_{fen}(N_{fen}-1)} & -\frac{(N_{fen}+1)}{N_{fen}(N_{fen}-1)} \end{bmatrix} \Delta X^\top \tag{3}$$

where:

$$\begin{cases} \delta X^+ = X^{new} - M_k^j(t-1), \\ \delta X^- = X^{old} - M_k^j(t-1), \\ \Delta X = [\delta X^+ \quad \delta X^-]. \end{cases} \tag{4}$$

with $M_k^j(t)$ and $\Omega_k^j(t)$ respectively center and covariance matrix of the class C_k^j at time t, N_{fen} the width of the sliding window, $X^{new} = X(t)$, X^{old} the old data in the set affected to C_k^j.

Then, the distances between the normal, current and default classes can be computed according to the center and the covariance matrix of each class. The Euclidean distance corresponds to the distance between the center of two classes:

$$d_{Eu} = (M_1^j - M_2^j)^\top (M_1^j - M_2^j) \tag{5}$$

where M_1^j and M_2^j are the centers of the classes C_1^j and C_2^j respectively.

The Kullback-Leibler distance corresponds to the distance between two classes taking into account their shape, *i.e.* the covariance matrices, (Kullback and Leibler, 1951). In the general case (Anguita and Hernando, 2004), the distance between the classes C_1^j and C_2^j is expressed by:

$$d_{kl}(C_1^j, C_2^j) = \frac{1}{2}(M_1^j - M_2^j)^\top (\Omega_1^{-1} + \Omega_2^{-1})(M_1^j - M_2^j) + \frac{1}{2}trace(\Omega_1^{-1}\Omega_2 + \Omega_1\Omega_2^{-1}) - d. \tag{6}$$

where d is the dimension of the data representation space, $\Omega_1 = \Omega_1^j$ and $\Omega_2 = \Omega_1^j$ are the covariance matrices of the classes C_1^j and C_2^j. The second term of d_{kl}, *i.e.* $trace()$, is specifically impacted by the shape and the orientation of the classes.

Finally, the distances between the several modes are used to estimate the probabilities of the failure occurrence of each component, as detailed in the next section.

4 ESTIMATION OF FAILURE OCCURRENCE PROBABILITIES

The probability of the failure occurrence, denoted $p^{E_j}(t)$, is defined as the PFE of an elementary component, and is considered as an indicator of the deterioration of this component. It is estimated according to the distance covered by the current class towards the default class, $\alpha(t)$, due to slow drifts:

$$p^{E_j}(t) = 1 - \alpha(t) \tag{7}$$

with:

$$\alpha(t) = \frac{distance(C_p^j, C_e^j(t))}{distance(C_n^j, C_p^j)} \tag{8}$$

where C_n^j, C_p^j, and C_e^j are the normal, default and current classes. The distance between two classes is computed according to the Euclidean (5) or the Kullback-Leibler (6) methods. It is assumed that $0 \leq \alpha(t) \leq 1$.

- **Estimation of $p^{E_j}(t)$ based on Euclidean Distance**

The percentage of distance $\alpha_{Eu}(t)$ which is estimated according to the Euclidean distance (5), is used to determine the probability $p_{Eu}^{E_j}(t)$ according to the relation (7). The example shown in Figure 3 is considered to illustrate this method. Three classes for component j are represented: normal C_n^j, current C_e^j, and default C_p^j classes characterized by (M_n^j, Ω_n^j), (M_e^j, Ω_e^j), and (M_p^j, Ω_p^j), respectively.

The percentage of distance $\alpha_{Eu}(t)$ at each time t is given by:

$$\alpha_{Eu}(t) = \frac{d_{Eu}(M_p^j, M_e'^j(t))}{d_{Eu}(M_n^j, M_p^j)} \tag{9}$$

where the distances d_{Eu} are expressed by relation (5), and $M_e'^j$ is the orthogonal projection of the center M_e^j on the segment $[M_n^j\, M_p^j]$. The distance $d_{Eu}(M_n^j, M_p^j) =$

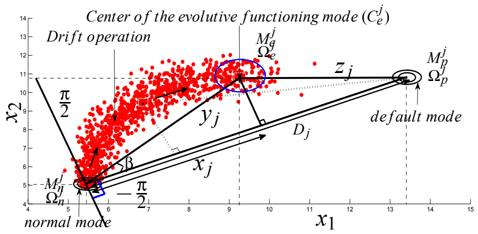

Figure 3: Evolution of a class from the normal class to the default class.

D_j is constant (10). The distance $d_{Eu}(M_n^j, M_e^{\prime j}(t)) = x_j$ is determined according to relation (11) from the triangle formed by the centers of the classes (*see* Figure 3). It is assumed that the current class can only evolve towards the default class. Consequently, the angle β is always include between $-\frac{\pi}{2} < \beta < \frac{\pi}{2}$, and the orthogonal projection of the center M_e^j is always included in the segment $[M_n^j\ M_p^j]$.

$$D_j = \sqrt{(M_n^j - M_p^j)^\top (M_n^j - M_p^j)} \tag{10}$$

$$x_j(t) = \frac{1}{2}\left[D_j + \frac{y_j^2(t) - z_j^2(t)}{D_j}\right] \tag{11}$$

with:

$$y_j(t) = \sqrt{(M_n^j - M_e^j(t))^\top (M_n^j - M_e^j(t))} \tag{12}$$

$$z_j(t) = \sqrt{(M_p^j - M_e^j(t))^\top (M_p^j - M_e^j(t))} \tag{13}$$

(14)

- **Estimation of $p^{E_j}(t)$ based on Kullback-Leibler Distance**

The Kullback-Leibler distance is used to estimate the percentage of distance $\alpha_{Kl}(t)$ and then to determine the probability $p_{Kl}^{E_j}(t)$, according to the relation (7). The percentage of distance $\alpha_{Kl}(t)$ at each time t is given by:

$$\alpha_{kl}(t) = \frac{d_{kl}(C_p^j, C_e^j(t))}{d_{kl}(C_n^j, C_p^j)} \tag{15}$$

where the distances d_{kl} are expressed by relation (6). The Kullback-Leibler distance between the class C_n^j and the class C_p^j is constant.

The percentage of distance $\alpha_{Kl}(t)$ is computed only when the current class C_e^j evolves towards the default class C_p^j. A criterion T_c^d is defined to verify this condition (16). Thus, $\alpha_{Kl}(t)$ is computed if and only if the criterion T_c^d is strictly negative.

$$\begin{cases} T_c^d = \dfrac{1}{N_{fen} - 1} \displaystyle\sum_{t=2}^{N_{fen}} sign(\Delta t), \\ \Delta(t) = d_{kl}(C_p^j, C_e^j(t)) - d_{kl}(C_p^j, C_e^j(t-1)) \end{cases} \tag{16}$$

- **Interpretation of $p^{E_j}(t)$ Computed According to Euclidean and Kullback-Liebler Distances**

The probabilities $p_{Eu}^{E_j}(t)$ and $p_{Kl}^{E_j}(t)$ are computed according to the Euclidean or Kullback-Liebler distances $\alpha_{Eu}(t)$ and $\alpha_{Kl}(t)$. To interprete and verify the pertinence of these indicators and thus the proposed methods, a scenario which consists in four current classes C_1^j to C_4^j which evolve to the normal class C_n^j towards the default class C_p^j, is considered and depicted in Figure 4. The classes C_1^j, C_2^j and C_3^j have the same centers but their matrices of covariance are different. The class C_4^j is characterized by different center and covariance matrice. The probabilities $p_{Eu}^{E_j}(t)$ and $p_{Kl}^{E_j}(t)$ are computed for the four classes. The results are given in Table 1.

Table 1: Probabilities computed for the classes.

	C_1^j	C_2^j	C_3^j	C_4^j
$p_{Eu}^{E_j}$	0,50	0,50	0,50	0,56
$p_{Kl}^{E_j}$	0,36	0,50	0,46	0,46

The Euclidean distance leads to the estimation of a same pourcentage $p_{Eu}^{E_j}$ for classes C_1^j, C_2^j and C_3^j, and to a pourcentage more important for the class C_4^j. Indeed, the center of the class C_4^j is nearest to the default class than the others classes C_p^j. This distance is easily interpretable but it does not take into account the shape and the orientation of the classes.

The Kullback-Liebler distance leads to the estimation of pourcentages $p_{Kl}^{E_j}$ different for the classes C_1^j, C_2^j and C_3^j. Although, the covariance matrix of the class C_1^j is smaller than the covariance matrix of the class C_3^j, the difference between the obtained pourcentages seems to be too important, and these indicators are not directly interpretable as the probabilities of the failure occurrence. Moreover, the pourcentages $p_{Kl}^{E_j}$ of the classes C_3^j and C_4^j are identical although the class C_4^j is nearest of the default class (*see* Figure 4). Finally, the Kullback-Liebler distance allows to take into account the shape and the orientation of the classes, but it is not directly usable for the estimation of the probabilities of the failure occurrence.

- **New Estimation Method of $p^{E_j}(t)$**

A new estimation method of the probability $p^{E_j}(t)$ is proposed to provide pertinent indicators which take into account in priority the position of the classes, but also, the remoteness, enlarging and rotation of these classes. It consists in a weigthed combination of $p_{Eu}^{E_j}(t)$ computed according to Euclidean distance and

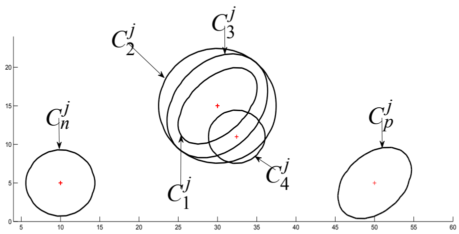

Figure 4: Scenario of evolution of classes from a normal class towards a default class.

p_ε which is computed according to the second term of the Kullback-Leibler distance (6). The probability $p^{E_j}(t)$ is expressed as:

$$p^{E_j}(t) = p_{Eu}^{E_j}(t) + \lambda\, p_\varepsilon \qquad (17)$$

where λ $(0 < \lambda < 1)$ is a weight coefficient. The parameter p_ε is function of the covariances matrices of the normal, default and current classes:

$$\begin{aligned} p_\varepsilon &= \frac{T_1}{T_1 + T_2} \\ T_1 &= trace(\Omega_e \Omega_p^{-1} + \Omega_e^{-1}\Omega_p) \\ T_2 &= trace(\Omega_p \Omega_n^{-1} + \Omega_p^{-1}\Omega_n) \end{aligned} \qquad (18)$$

The coefficient λ is tuned in order to take into account the covariance matrices in the estimation of $p^{E_j}(t)$ without however obtaining too important differences between the distances from the classes. In Table 2, we presente the occurrence probabilities computed by relation (17) according to $\lambda = 1/10$.

Table 2: Failure occurrence probabilities.

	C_1^j	C_2^j	C_3^j	C_4^j
$p^{E_j}(t)$	0,53	0,55	0,54	0,60

If the value of λ is too small, the shape of the class is not taken into account, and that leads at considering only the Euclidean distance. If the value of λ is too big, the shape of the class has too much influence on the estimation of $p^{E_j}(t)$, and that leads to the same problem of interpretation than the distance of Kullback-Leibler. The proposed method is applied on a real scenario in the next section.

5 APPLICATION

A temperature controller is a process which is used to control the temperature of a client system. It is composed of an electric heater, a pump, a heat exchanger and a filter (*see* Figure 5). The components of this

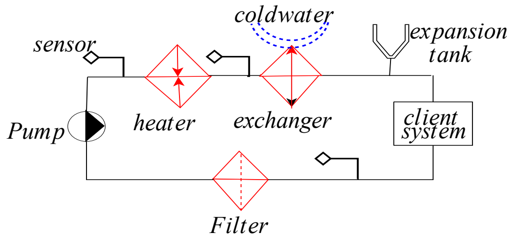

Figure 5: Thermo-regulator components.

process are subject to failures related to slow degradations due to scaling and fouling essentially. If these failures are not taken into account early enough, they can cause the stop of the process.

The first step is the FMEA of the temperature controller which allows the determination of the Fault Tree of the process (*see* Figure 6). The Fault Tree is composed of three basic events associated to each component and a top event which correspond to the no temperature control. The basic events are:

- Failure of the heater (E_1)
- Failure of the exchanger (E_2)
- Failure of the filter (E_3)

the top event is:

- No temperature control (E_4)

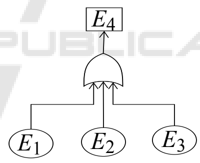

Figure 6: Fault Tree of the temperature controller.

The temperature controller is equipped by sensors located at the input and output of each component. These sensors measure the pressure of the fluid. An observation vector is done by $X_1 = (x_1, x_2, x_3)^\top$ where the three indicators are determined according to the measurements:

$$x_1 = \frac{P_{input\ heater} - P_{output\ heater}}{\Delta P_{pump}} \qquad (19)$$

$$x_2 = \frac{P_{input\ exchanger} - P_{output\ exchanger}}{\Delta P_{pump}} \qquad (20)$$

$$x_3 = \frac{P_{input\ filter} - P_{output\ pump}}{\Delta P_{pump}} \qquad (21)$$

where x_1, x_2, x_3 are indicators to monitor the heater, the exchanger and the filter respectively.

The AUDyC technique allows the monitoring of elementary components of the temperature controller.

The estimation method of the occurrence probabilities, with a weight coefficient tuned as $\lambda = 1/20$, $\lambda = 1/50$, is used to estimate in real time the $p^{E_j}(t)_{j=1,2,3}$ (17) of each elementary component, and finally the PFE of top event (E_4) by propagation the basic events.

where the events E_1, E_2 and E_3 are independents. Thus, the PFE of the event E_4 is expressed as:

$$PFE(E_4) = ((p^{E_4}(t_1), t_1), \cdots, (p^{E_4}(t_n), t_n)) \quad (22)$$

In the real scenario considered, the components of the temperature controller are subjected to drifts as depicted in Figure 7. The $PFE(E_4)$ determined according to the relation (22) are displayed in Figure 8. On this real scenario,the tune $\lambda = 1/20$ leads to a too important influence of p_ε, whereas $\lambda = 1/50$ presents a good compromise. La figure 8.a montre l'influence de la forme de la classe alors que la figure 8.b l'influence de la forme de la classe est moins important.

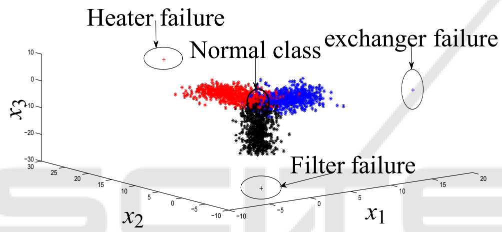

Figure 7: Drifts of operating of the components.

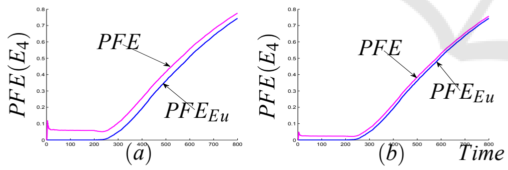

Figure 8: $PFE(E_4)$ according to (a) $\lambda = 1/20$, (b) $\lambda = 1/50$.

6 CONCLUSIONS

The supervision method proposed in this paper allows the estimation of the probability of failure occurrence of processes in real time. The dynamic clustering method is used to track the evolution of operating modes of processes by determining the characteristics of each class (center and covariance matrix).

The center and the covariance matrix being adapted by AUDyC, the Euclidean distance and trace of the covariance matrices are used to estimate the probability of the failure occurrence. The Euclidean distance does not allow to take into account the shape and the orientation of the class, and the Kullback-Leibler distance, are not easily interpretable. Then, a new method which is based on the weight combination between the probabilities estimated with the Euclidean distance and with the trace of the covariance matrices, is proposed and illustrated on real case. In futur works, we will propose a prognosis strategy based on this method to forecast the occurrence probability of events, and a step to tune the weight coefficients of the proposed method. The goal is to determine indicators to improve the predictive maintenance of processes. This will be implemented for predictive maintenance of the temperature controller and of measure the apport of the proposed methods.

REFERENCES

Anguita, J. and Hernando, J. (2004). Inter-phone and inter-word distances for confusability prediction in speech recognition. *Congreso de la Sociedad Espaola para el Procesamiento del Lenguaje Natural*, (33):33–40.

Desinde, M., Flaus, J. M., and Ploix, S. (2006). Tool and methodology for online risk assessement of process. In *Lambda-Mu 15 /Lille*.

Grall, A., Berenguer, C., and Dieulle, L. (2002). A condition-based maintenance policy for stochastically deteriorating systems. *Reliability Engineering and System Safety*, 76(2):167–180.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annal of Mathematical Statistics*,22:79-86.

Lassagne, M. (2000). Applying a decision-analysis-based method to the evaluation of potential risk-reducing measures : The case of a floating production storage and offloading unit in the gulf of mexico. *SPE annual technical conference, Dallas TX , USA*.

Lecoeuche, S., Lurette, C., and Lalot, S. (2004). New supervision architecture based on on-line modelling of non-stationary data. *Neural Computing and Applications Journal*, 13:323–338.

Lurette, C. and Lecoeuche, S. (2003). Unsupervised and auto-adaptive neural architecture for on-line monitoring. application to a hydraulic process. *Engineering Applications of Artificial Intelligence*, 16:441–451.

Mouchawed, S. M. and Billaudel, P. (2002). Influence of the choice of histogram parameters at fuzzy pattern matching performance, int. journal of wseas transactions on system. *WSEAS Transactions on Systems*, 1:260–266.

Muller, A., Suhner, M.-C., Iung, B., and Morel, G. (2004). Prognosis-based maintenance decision-making for industrial process performance optimisation. *In 7th IFAC Symposium on Cost Oriented Automation (COA2004). Gatineau/Ottawa Canada*.

Vesely, W. E., Goldberg, F. F., Robert, N. H., and Haasl, D. F. (1981). *Fault Tree Handbook*. US nuclear Regulatory Commission, Washington D.C., USA.