PERFORMANCES THEORETICAL MODEL-BASED OPTIMIZATION FOR INCIPIENT FAULT DETECTION WITH KL DIVERGENCE

Abdulrahman Youssef[⋆†] *Claude Delpha*[⋆] *Demba Diallo*[†]

[⋆] Laboratoire des Signaux et Systemes (L2S), CNRS, Supelec, Univ. Paris-Sud
[†]Laboratoire de Genie Electrique de Paris (LGEP), CNRS, Supelec, Univ. P. et M. Curie, Univ. Paris-Sud
91192 Gif Sur Yvette, France

ABSTRACT

Sensible and reliable incipient fault detection methods are major concerns in industrial processes. The Kullback Leibler Divergence (KLD) has proven to be particularly efficient. However, the performance of the technique is highly dependent on the detection threshold and the Signal to Noise Ratio (SNR). In this paper, we develop an analytical model of the fault detection performances (False Alarm Probability and Miss Detection Probability) based on the KLD including the noisy environment characteristics. Thanks to this model, an optimization procedure is applied to set the optimal fault detection threshold depending on the SNR and the fault severity.

***Index Terms*—** Fault detection, performance modeling, Optimization, Kullback-Leibler Divergence, Principal Component Analysis.

1. INTRODUCTION

Fault detection plays a key role in enhancing today's technological systems high demands for performance and security. In such systems, minor faults can result in catastrophic consequences, where the huge need for very sensitive fault detection and diagnosis (FDD) methods. Such methods must be insensitive to the environment evolution (noise, temperature,...) but also to input changes. In the opposite, they have to be very sensitive to the fault severity [1]. For industrial process monitoring, it is crucial to be able to detect very small faults (namely incipient modification) without stopping the process. Thus, it allows to prevent the subsequent occurrence of more catastrophic events. However, working with these incipient faults, leads to a big difficulty in distinguishing the fault itself from noise or environmental changes. Indeed, this can affect the performance of the detection method in term of false alarm probability (reliability) and miss alarm probability (sensitivity).

Process-history based methods are very commonly used for fault detection and diagnosis, which do not assume any form of physical model structure [2]. They rely only on process historical data. However, faults occurring in a process can be effectively detected and diagnosed with only few variables. Since each variable characterises a fault in a different manner, one variable will be more sensitive to certain faults and less sensitive to other ones. This motivates using multiple process monitoring variables, with the proficiency of each variable determined for the particular process and the possible faults at hand.

With a large amount of variables measured and stored automatically, multivariate statistical monitoring methods have become increasingly common in process industry. Specifically, Principal Component Analysis (PCA)-based process monitoring methods have gained wide application in industries. PCA-based monitoring methods can easily handle high dimensional, noisy and highly correlated data generated from industrial processes, and provide superior performance compared to univariate methods. In addition, these process monitoring methods are attractive because they only require a good historical data set of healthy operation, which are easily obtainable for computer-controlled industrial processes. The PCA can be used to reduce the m dimensional space of process variables to a lower l -dimensional subspace termed the principal subspace [3].

The most common procedure of process monitoring by PCA is to use T^2 and Q statistics to detect process faults exhibited in the latent variables and the residuals. The Kullback-Leibler Divergence (KLD) has been shown as alternative to the T^2 and SPE criteria for the detection of incipient faults under the PCA framework [4].This measure has been also used for abnormality detection and pattern recognition in different areas. Compared to existing work in the literature, it has been shown that the monitoring strategy with KLD under PCA is conceptually more straightforward and it is also more sensitive in detecting incipient fault [4], [5]. Indeed, the detection of this kind of faults is more complex with nuisance parameters [6]. In this paper, we propose to evaluate the performance of the KL divergence in such environment.

The performances of the method are dependent on the internal parameters (named as hyperparameters) but also on the process environment [7]. We propose here to develop a theoretical model of the performances. Then a deterministic optimization algorithm is used to obtain the optimized detection according to the fault severity and noise level.

2. INCIPIENT FAULT DETECTION METHOD

2.1. Notation

Let's introduce the following notations:
Let us set $X_{[N\times m]}$ such as $X = (x_1, ..., x_j, ..., x_m) = (x_{ij})_{i,j}$ is the original data matrix where $x_j = [x_{1j}...x_{Nj}]'$ is a column vector of N measurements taken for the jth variable.
$\bar{X}_{[N\times m]}$, where $\bar{X} = (\bar{x}_1, ..., \bar{x}_j, ..., \bar{x}_m)$ is the centered matrix; each column of X is subtracted from its mean value.
S is the sample data covariance matrix.
$P_{[m\times m]}$, such as $P = (p_1, ..., p_l, ...p_m)$ is the loading eigenvectors matrix.
$T_{[N\times m]}$, where $T = (t_1, ..., t_l, ..., t_m)$ is the scores matrix given by $T = \bar{X}P$
l is the dimension of the principal subspace and the number of latent scores as well.
$\lambda_1, ..., \lambda_l$ in the descendant order, are the variances of the latent scores and the eigenvalues associated respectively to $p_1, ..., p_l$.
a is the fault amplitude parameter.
$v_{[N\times m]}$ is the noise matrix .
The star mark (*) refers to the healthy and noise-free case.

2.2. Fault detection procedure reminder

The general procedure of statistical monitoring is to collect a large number of healthy data samples used as reference data set. The measured data are then compared to the healthy ones to check whether abnormal conditions occur.
Since principal components are concerned with variances, the information supplied by the data lies within the distributions along the principal components axes. Therefore, monitoring probability distributions of the latent scores, which represent the data projection onto the l principal components, will be able to reveal small changes caused by incipient faults. So, once the PCA's model is established, a reference probability distribution is estimated for each latent score. Then for each new set of observations, the associated latent scores are calculated through the PCA's model. A natural idea consists in measuring the difference between the probability density functions of healthy data and measured ones.

2.2.1. Kullback-Leibler Divergence for Detection

The difference between the probability density functions of healthy data and test data can be achieved by the KLD computation between the two distributions [8].
For discrimination between two continuous probability density functions (pdfs) $f(x)$ and $g(x)$ of a random variable x, the Kullback-Leibler Information (KLI) is defined as:

$$I(f\|g) = \int f(x) \log \frac{f(x)}{g(x)} dx. \tag{1}$$

The KL Divergence (KLD) is then defined as the symmetric version of the KL Information [8], [9]:

$$KLD(f, g) = I(f\|g) + I(g\|f). \tag{2}$$

2.2.2. KLD approximation by Monte Carlo Simulation

For arbitrary distributions f and g, (1) can be numerically approximated using Monte Carlo (MC) simulation. The Monte Carlo method expresses (1) as the expectation of $log(f/g)$, under the pdf f. Using n_s i.i.d samples $\{z_i\}_1^{n_s}$ drawn from f, it consists in calculating

$$D_{MC}(f, g) = \frac{1}{n_s} \sum_{i=1}^{n_s} log \frac{f(z_i)}{g(z_i)} \tag{3}$$

The estimation error is normal with variance σ^2_{MC} and zero mean ($\sim \mathcal{N}(0, \sigma^2_{MC})$) such as $\sigma^2_{MC} = \frac{1}{n_s} VAR_f[log(f/g)]$.

2.3. KLD theoretical model for fault detection

For normal densities f and g such that $f \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $g \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where μ_1, μ_2 are the means and σ_1^2, σ_2^2 are the variances for f and g respectively, the Kullback-Leibler Divergence between f and g is given by:

$$KLD\,(f, g) = \frac{1}{2}\left[\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + (\mu_1 - \mu_2)^2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) - 2\right] \tag{4}$$

A simple and light computational expression of the divergence is obtained assuming that the measurements vector $X = [x_1, x_2, ..., x_m]$ is $m-$variate normally distributed. So principal component scores, which are linear combinations of the original variables, are also normally distributed. However, by contrast to the last $(m - l)$ residual scores (the data projection onto the residual subspace), the latent ones have large variances so that their distributions are far from being degenerated. Therefore, the divergence is strongly related to the principal subspace.
From the assumption of normality, it follows that each of the l principal scores $t_k, (k = \{1, 2, ..., l\})$, has a pdf which we denote f such that $f \sim \mathcal{N}(0, \lambda_k)$. We propose to compare f against its reference. The reference is denoted f^*, $f^* \sim \mathcal{N}(0, \lambda_k^*)$. It is totally described by the eigenvalue λ_k^* which refers to the PCA's model. The mean of the distribution is supposed unchanged (zero) after the noise and fault occurrence, because we assume that a noise and a fault, particularly an incipient one, will not move the centre of the PCA's model. This assumption has been made with the detection of subspace changes approach. Then, we can write:

$$\lambda_k = \lambda_k^* + \Delta\lambda_k^a + \Delta\lambda_k^v \tag{5}$$

where $\Delta\lambda_k^a$ is the eigenvalue bias caused by the fault occurrence. The fault affecting the jth variable x_j among the m process variables is considered as an additive bias of amplitude on this variable with factor a, occurring within the sampling interval $[b, c]$.

and $\Delta\lambda_k^v$ is the eigenvalue bias caused by the noise presence. By specializing (4) to the case considered, the divergence becomes:

$$KLD\left(f,f^*\right)=\frac{1}{2}\left[\frac{\left(\Delta\lambda_k^a+\Delta\lambda_k^v\right)^2}{\lambda_k^*\left(\lambda_k^*+\Delta\lambda_k^a+\Delta\lambda_k^v\right)}\right] \tag{6}$$

The fault is characterized by its amplitude a, the noise by the vector v_k. p_k^* is the kth loading eigenvector associated to λ_k^*, they refer to the PCA's model for which $a=0$ and $v_k=0$.

From (6), the theoretical model of the KLD can be formulated. Finally, the theoretical expression of the divergence between the pdf of the kth principal score and its reference, depending on the fault amplitude parameter a is hence given, from (6), as (7). In (7) δ_r, γ_v, δ_j and σ are constants independent of the fault parameter. δ_r, δ_j and σ are given in function of the original variables, which are not faulty and not noisy, and the fault-free measurements of the variable x_j as well, and γ_v is given in function of the original variable x_j and the noise v. The computation of these constants requires however the knowledge of the faulty interval $[b,c]$. The details are given in [6].
The eigenvalue bias caused by the noise presence $var(v\times p_k^*)$ can be expressed in function of the Signal to Noise Ratio (SNR) as:

$$var\left(v\times p_k^*\right)=\left(\sum_{i=1}^{m}p_{ik}\right)^2*\frac{P_s}{10^{(SNR/10)}} \tag{8}$$

Where P_s is the signal power.
Moreover, the variable γ_v is equal to zero if we suppose that the noise present in the system is an additive white gaussian noise ($\gamma_v=\sum_{i=b}^{c}x_{ij}^*v_{ij}-\frac{1}{N}\sum_{i=b}^{c}x_{ij}^*\sum_{i=1}^{N}v_{ij}$) [6]. Then, substituting $var(v\times p_k^*)$ by it's expression in (7) and replacing γ_v by 0, the KLD model becomes (9).

With the healthy case ($a=0$), the KLD model (9) becomes:

$$KLD_{M0}\left(f,f^*\right)=\frac{1}{2}\frac{\left(\left(\sum_{i=1}^{m}p_{ik}\right)^2*\frac{P_s}{10^{(SNR/10)}}\right)^2}{\lambda_k^*\left(\lambda_k^*+\left(\sum_{i=1}^{m}p_{ik}\right)^2*\frac{P_s}{10^{(SNR/10)}}\right)} \tag{10}$$

To detect the fault occurrence, the value of the KLD is compared to a threshold h. If the KLD surpasses h, an alarm is signaled. In the past work, the choice of h is set to $h=KLD_{M0}+\alpha*\sigma_{MC0}$ with $\alpha=2$ arbitrarily fixed. The goal of the following work is to find the optimal detection threshold (optimal α value) in a noisy environment.

3. PERFORMANCE ANALYSIS

3.1. Performance Modeling

A key issue in fault detection method is to state the significance of the observed deviation (fault) with respect to random noises, deterministic uncertainties (also called nuisance parameters [1]. A main challenge of the statistical methods is their ability to handle noises and uncertainties, to reject nuisance parameters, to decide between two hypothesis H_0 (no faults $a=0$) and H_1 (there exists a fault $a\neq 0$).
The performance of the hypothesis test is characterized by False Alarm Probability (P_{FA}) and the Miss Detection Probability (P_{MD}). These criteria depend on the test threshold h, the noise level in the system and the fault severity.

$$P_{FA}=\mathcal{P}(KLD>h\mid H_0) \tag{11}$$

$$P_{MD}=\mathcal{P}(KLD<h\mid H_1) \tag{12}$$

To calculate the false alarm and miss detection probabilities, the law of the KLD variation should be known. However, the estimation of the KLD by the Monte carlo simulation assumes that the KLD has a gaussian variation. To validate this assumption, the Kolmogorov-Smirnov test is applied on a set of KLD calculated for the same noise and fault levels. The test results confirm the assumption that the KLD has a gaussian variation (confidence limit 95%). Then $KLD\sim\mathcal{N}(KLD_M,\sigma_{mc}^2)$, where KLD_M is the value of the theoretical model and σ_{mc}^2 is the variance of the Monte Carlo estimation. The probability density function (pdf) of the KLD is then:

$$f(x)=\frac{1}{\sigma_{mc}\sqrt{2\pi}}e^{-\frac{(x-KLD_M)^2}{2\sigma_{mc}^2}} \tag{13}$$

The probability density function of the KLD under H_0 is $f_0(x)$ where $KLD_M=KLD_{M0}$ is the theoretical value in the fault-free or healthy case ($a=0$) and $\sigma_{mc}=\sigma_{mc0}$ is the corresponding standard deviation. Under H_1 the KLD Pdf is $f_1(x)$ where $KLD_M=KLD_{M1}$ is the theoretical value in the faulty case ($a\neq 0$) and $\sigma_{mc}=\sigma_{mc1}$ is the corresponding standard deviation.
Now, from (13), the expressions of the False Alarm Probability (P_{FA}) and Miss Detection Probability (P_{MD}) are:

$$P_{FA}=\mathcal{P}(KLD>h\mid H_0)=1-\int_{-\infty}^{h}f_0(x)dx \tag{14}$$

$$P_{MD}=\mathcal{P}(KLD<h\mid H_1)=\int_{-\infty}^{h}f_1(x)dx \tag{15}$$

However, the cumulative distribution function of the gaussian law doesn't exist, so to calculate P_{FA} and P_{MD}, the Taylor approximation is used:

$$f(x)\simeq R(x,c,n)=\sum_{n=0}^{N_t}\frac{f^{(n)}(c)}{n!}(x-c)^n \tag{16}$$

The methodology to calculate P_{FA} and P_{MD} is:

1. Choose the lower limit ξ such that $f(x)=0$, if $x\leq\xi$.
2. Create a partition C with β elements in the interval $[\xi,x]$ such as $C=\{c_1,c_2,...,c_\beta\}$ with step $2d=\frac{x-\xi}{\beta}$.
3. Choose the Taylor function order N_t

$$KLD_M(f,f^*) = \frac{2}{N^2}\frac{\left(p_{jk}\sum_{r=1}^{m}\left(\delta_r+\gamma_v\right)p_{rk}\times a+\frac{1}{2}p_{jk}^2\sigma\times a^2+\frac{N}{2}var\left(v\times p_k^*\right)\right)^2}{\lambda_k^*\left(\lambda_k^*+\frac{2}{N}\left(p_{jk}\sum_{r=1}^{m}\left(\delta_r+\gamma_v\right)p_{rk}\right)\times a+\frac{1}{N}p_{jk}^2\sigma\times a^2+var\left(v\times p_k^*\right)\right)} \quad (7)$$

$$KLD_M\left(f,f^*\right) = \frac{2}{N^2}\frac{\left(p_{jk}\sum_{r=1}^{m}\delta_r p_{rk}\times a+\frac{1}{2}p_{jk}^2\sigma\times a^2+\frac{N}{2}\left(\sum_{i=1}^{m}p_{ik}\right)^2 * \frac{P_s}{10^{(SNR/10)}}\right)^2}{\lambda_k^*\left(\lambda_k^*+\frac{2}{N}\left(p_{jk}\sum_{r=1}^{m}\delta_r p_{rk}\right)\times a+\frac{1}{N}p_{jk}^2\sigma\times a^2+\left(\sum_{i=1}^{m}p_{ik}\right)^2 * \frac{P_s}{10^{(SNR/10)}}\right)} \quad (9)$$

Thus, applying this methodology, we obtain:

$$\Longrightarrow \int_{c_i-d}^{c_i+d} f(x)dx \simeq \int_{c_i-d}^{c_i+d} R(x,c_i,N_t)dx$$

$$= f(c_i)(x-c_i) + \frac{f'(c_i)}{2}(x-c_i)^2 + ...$$

$$... + \frac{f^{(N_t+1)}(c_i)}{(N_t+1)!}(x-c_i)^{N_t+1}\,|_{c_i-d}^{c_i+d}$$

$$= f(c_i)(2d) + \frac{f''(c_i)}{6}(2d)^3 + ... + \frac{f^{(N_t)}(c_i)}{(N_t+1)!}(2d)^{N_t+1}$$

$$\Longrightarrow \int_{-\infty}^{x} f(x)dx \simeq \sum_{i=1}^{\beta}\sum_{2j+1=0}^{N_t}\frac{f^{(2j)}(c_i)}{(2j+1)!}(2d)^{2j+1}$$

Choosing $N_t = 3$, we obtain:

$$P_{FA} = 1 - \sum_{i=1}^{\beta}\left(f_0\left(c_i\right).\frac{h-\xi}{\beta} + \frac{f_0''\left(c_i\right)}{6}.\left(\frac{h-\xi}{\beta}\right)^3\right) \quad (17)$$

$$P_{MD} = \sum_{i=1}^{\beta}\left(f_1\left(c_i\right).\frac{h-\xi}{\beta} + \frac{f_1''\left(c_i\right)}{6}.\left(\frac{h-\xi}{\beta}\right)^3\right) \quad (18)$$

Substituting f_0 and f_1 by their expressions, we obtain the model of the performance with respect to threshold, fault severity and SNR in (19) and (20). Note that the fault severity and SNR are included in the KLD_M expression.

Setting a high threshold improves the P_{FA} but degrades the P_{MD} and setting a low value improves the P_{MD} and degrades the P_{FA}. Therefore, both criteria are combined to design a COST function that will be optimized.
To combine the two criteria, we define the cost function such as $COST = P_{FA} + P_{MD}$ [10]. The variation of the COST with respect to the threshold factor α such as $h = KLD_{M0} + \alpha * \sigma_{MC0}$ is studied according to the fault severity a and the environment noise influence. To show the effect of the noise level on the detection capability, it has to be compared to the fault amplitude, this is done using the Fault to Noise Ratio (FNR) [4], [6] defined as $FNR = 10log(\frac{P_f}{P_v})$. Where P_f is the fault power and P_v is the noise power. The COST variation with α and FNR is obtained from the theoretical performance model and plotted in Fig.1. Note that these results are obtained using $\beta = 6$ in the COST function.

As seen in Fig.1, the performances of the fault detection methods depend strongly on the choice of the threshold detection.

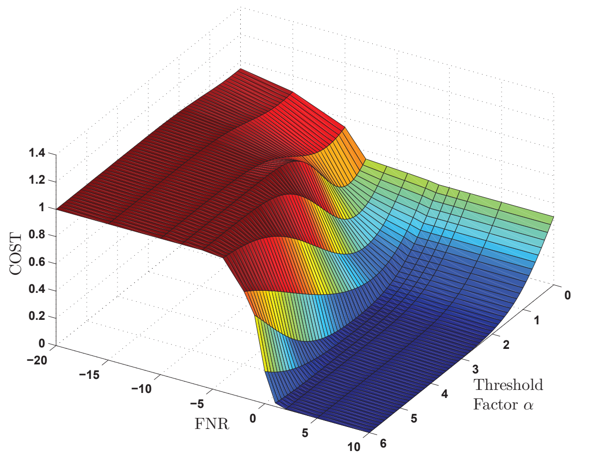

Fig. 1. COST variation along with FNR and Threshold factor

3.2. Performance Optimization

Searching the best tuning of the detection threshold is formulated as an optimization problem, where the cost function is that obtained in the previous section. Since we have a theoretical model of the cost function, a deterministic algorithm of optimization can be applied. In this paper, as the COST function has not several local minima, the descent gradient method is suitable.
Gradient descent is an algorithm for finding the nearest local minima of a function which presupposes that the gradient of the function can be computed. In Fig.2, the progress of the optimization algorithm is displayed as an example for 3 different FNR (FNR=[0dB, -3dB, -6dB]), showing that for each FNR an optimum α value can be found. Fig.3 shows the evolution of the optimal threshold factor α for each FNR with the associated COST: the α value increases along with the FNR value but becomes quite constant for FNR values higher than 1dB. In the opposite the COST function decreases along with the FNR and becomes almost null for FNR higher than 1dB. As it can be seen in Fig.3, if FNR=0dB, the usual setting of $\alpha = 2$ is justified. However for lower values of the FNR, one can see that a suitable optimized threshold value can be obtained to guarantee the best compromise between P_{FA} and P_{MD}. Therefore, with the FDD technique and the analytical model, the reliability and the sensitivity can be tuned optimally.

$$P_{FA} = 1 - \sum_{i=1}^{\beta}\left(\frac{1}{\sigma_{mc0}\sqrt{2\pi}}e^{-\frac{(c_i - KLD_{M0})^2}{2\sigma_{mc0}^2}}.\frac{h-\xi}{\beta} + \frac{-1}{\sigma_{mc0}^2}\left(1 - \frac{(c_i - KLD_{M0})^2}{\sigma_{mc0}^2}\right).\frac{1}{\sigma_{mc0}\sqrt{2\pi}}e^{-\frac{(c_i - KLD_{M0})^2}{2\sigma_{mc0}^2}}.\left(\frac{h-\xi}{\beta}\right)^3\right) \quad (19)$$

$$P_{MD} = \sum_{i=1}^{\beta}\left(\frac{1}{\sigma_{mc1}\sqrt{2\pi}}e^{-\frac{(c_i - KLD_{M1})^2}{2\sigma_{mc1}^2}}.\frac{h-\xi}{\beta} + \frac{-1}{\sigma_{mc1}^2}\left(1 - \frac{(c_i - KLD_{M1})^2}{\sigma_{mc1}^2}\right).\frac{1}{\sigma_{mc1}\sqrt{2\pi}}e^{-\frac{(c_i - KLD_{M1})^2}{2\sigma_{mc1}^2}}.\left(\frac{h-\xi}{\beta}\right)^3\right) \quad (20)$$

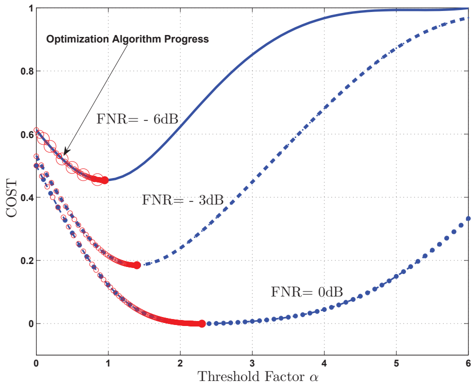

Fig. 2. Descent gradient method progress

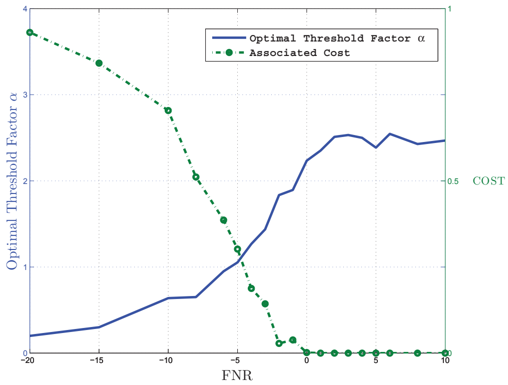

Fig. 3. Optimal threshold factor α and COST variation

4. CONCLUSION

This paper described the application of an optimal tuning methodology of the detection threshold for the Kullback-Leibler Divergence (KLD) method. The idea is to consider this problem as an optimization problem with a cost function defined as the performance response to a choice of detection threshold and environment conditions. A theoretical evaluation of the KLD allows establishing a performance theoretical model. Then, a deterministic optimization algorithm (descent gradient) is applied. The results of the optimization show that an optimal tuning of the detection threshold can be obtained to improve the performances of the detection method even while the environment noise level is higher than the fault severity power.

REFERENCES

[1] M. Fouladirad and I. Nikiforov, "Optimal statistical fault detetction with nuisance parameters," *Automatica*, vol. 41, no. 7, pp. 1157–1171, 2005.

[2] M. Basseville and I. Nikiforov, ," in *Detection of Abrupt Changes-Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ., 1993.

[3] I.T. Jollife, ," in *Principle Component Analysis*. Second Edition, Aberdeen U.K, 2002.

[4] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on kullback-leibler divergence using principal component analysis: Part I," *Elsevier Journal on Signal Processing*, vol. 94, pp. 278–287, January 2014.

[5] J. Harmouche, C. Delpha, and D. Diallo, "A theoretical approach for incipient fault severity assessment using the kullback-leibler divergence," in *EUSIPCO 2013*, Marrakech, Marocco, September 9-13 2013.

[6] A. Youssef, J. Harmouche, C. Delpha, and D. Diallo, "Capability evaluation of incipient fault detection in noisy environment : A theoretical kullback-leibler divergence-based approach for diagnosis," in *IEEE IECON 2013*, Vienna, Austria, November 10-13 2013.

[7] J. Marzat, E. Walter, F. Damongeot, and H.P. Lahanier, "Robust automatic tuning of diagnosis methods via an efficient use of costly simulations," in *IFAC*, Brussels, Belgium, 2012.

[8] S. Kullback and R.A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[9] M. Basseville, "Distances measures for signal processing and pattern recognition," *Elsevier Signal Processing*, vol. 18, no. 4, pp. 349–369, 1989.

[10] M. Bartys, R.J. Patton, M. Syfert, S. De las Heras, and J. Quevedo, "Introduction to the damadics actuator fdi benchmark study," *Control engineering practice*, vol. 14, no. 6, pp. 577–596, 2006.