ENHANCEMENT OF INCIPIENT FAULT DETECTION AND ESTIMATION USING THE MULTIVARIATE KULLBACK-LEIBLER DIVERGENCE

Abdulrahman YOUSSEF, Claude DELPHA

Laboratoire des Signaux et Systèmes,
CNRS, CentraleSupélec, Univ. Paris-Sud,
Université Paris-Saclay,
91192 Gif Sur Yvette, France

Demba DIALLO

Group of Electrical Engineering – Paris,
CNRS, CentraleSupélec, Univ. Paris-Sud,
Université Paris-Saclay,
Sorbonne Universités, UPMC Univ. Paris 06
91192 Gif Sur Yvette, France

ABSTRACT

Fault detection and diagnosis methods have to deal with large variable data sets encountered in complex industrial systems. Solutions to this problem require multivariate statistics approaches often focused on the reduction of the space dimension. In this paper we propose a fault detection and estimation approach using Multivariate Kullback-Leibler Divergence (MKLD) to cope with the negative effects due dimension reduction while using Principal Component Analysis (PCA). The obtained results show its superiority on the usual PCA-KLD based approach. An analytical model of the MKLD is proposed and validated for low severity fault (incipient fault) detection and estimation in noisy environment operating conditions.

***Index Terms*—** Incipient fault diagnosis, Detection, Estimation, Kullback-Leibler Divergence, Multivariate analysis

1. INTRODUCTION

The last three decades have shown an increased demand for improving the economy and safety of processes. Health monitoring of processes has been widely developed with studies on fault detection and diagnosis. In a wide variety of industrial and onboard applications, the detection and diagnosis of faults are considered essential to ensure high performance level of the plant operation to reduce economic losses and enhance the system security [1]. Due to increasing safety rules, and in order to avoid systems unwanted stops it is crucial to be able to detect at their early stage failures that will be able to occur disease in a complex system. Such incipient faults are difficult to detect due to their low severities and their high sensitivity to noisy environments.

It has been recognized that statistical-based techniques have many attractive advantages in dealing with large variable sets encountered in complex industrial systems. Among various statistical-based techniques, multivariate projection-based methods are the most popular ones [2]. Principal Component Analysis (PCA) is one of the most often used for multivariate data-driven-based industrial systems health monitoring [3, 4, 5]. Its main interest is the ability to reduce the data dimensional-space while keeping the maximum variance information available [6].

For fault detection purpose, statistical-based criteria in PCA framework have been successfully used. For example T^2, Q statistics, and $f-divergences$ techniques have shown their efficiency [7, 8, 9, 10]. Comparing those techniques, J. Harmouche et al [10, 11], have shown that the monitoring strategy with Kullback-Leibler Divergences (KLD) using PCA is conceptually more straightforward and also more sensitive for the detection of faults with very low severities namely incipient faults.

In [12], a PCA-KLD univariate-based approach focused on the maximum variance components has been used to detect incipient faults. It has been highlighted that the fault detection threshold, and the performances defined by the Missed Detection Probabilities (P_{MD}) and False Alarm Probabilities (P_{FA}) are strongly related to the Fault to Noise Ratio (FNR). Moreover, the incipient fault detections performances can be degraded by the projection error and the uncorrelated variables that can be yield using PCA [13]. To cope with this problem, B. Wang et al [14] propose to base their statistical analysis on a multi-block principal component analysis. In their approach the correlated signals with close similarities are first grouped in the same-block for the PCA. Detection analysis is done between the Multi-block PCA results. This solution is efficient but in the case of incipient fault with nuisance parameters (noise), this is not sufficient : the smaller the fault is, more difficult is the accurate fault detection and severity estimation. The performances are then limited due to loss brought by the PCA. Therefore, we propose here a multivariate KLD (MKLD) approach to evaluate the dissimilarities among measured data distribution without any PCA transformation so as to use the whole data information available for the analysis. Within this multivariate analysis, we propose then to estimate more accurately the fault severity in the noisy environment and reach better performances.

2. PAPER CONTRIBUTION

In this paper, considering a multivariate process history based method, we propose first to improve (reduce) the fault detectability of incipient fault using multivariate KLD in the original data framework. We will show how the P_{MD} and P_{FA} performances are improved and how the detection threshold (via the Fault to Noise Ratio) is minimized. Then we propose to estimate accurately the incipient fault severity with an analytical model of the multivariate KLD in a fault diagnosis context. The robustness of the estimated severity information for different noise levels are assessed.

3. INCIPIENT FAULT DETECTION

3.1. Kullback-Leibler Divergence

In multivariate statistical process monitoring, an important task is to monitor the divergence between healthy data and measured ones to first detect and then diagnose a fault occurrence. A natural idea

978-0-9928-6265-7/16/$31.00

consists of measuring the divergence between the probability density functions (Pdf) of healthy data and measured ones. This can be achieved by observing the Kullback-Leibler Divergence between the two probability distributions.

For discrimination purpose between two continuous Pdf $f(x)$ and $g(x)$ of a random variable x, the Kullback-Leibler Information (KLI) [15] is defined by $I(f\|g)$ such as $I(f\|g) = \int f(x)\log\frac{f(x)}{g(x)}dx$. The KLD is then defined as the symmetric version of the KL Information [16] denoted by $KLD(f,g) = I(f\|g) + I(g\|f)$.

For normal densities f and g such that $f \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and $g \sim \mathcal{N}(\mu_1, \sigma_1^2)$, where μ_0, μ_1 are the means and σ_0^2, σ_1^2 are the variances values for f and g respectively, the symmetrical univariate KLD between f and g is:

$$KLD(f,g) = \frac{1}{2}[\frac{\sigma_1^2}{\sigma_0^2} + \frac{\sigma_0^2}{\sigma_1^2} + (\mu_0 - \mu_1)^2(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2}) - 2] \quad (1)$$

The non-symmetrical Multivariate Kullback-Leibler Divergence (MKLD) [17] for Gaussian signal densities $f \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $g \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, with non-singular covariances matrices denoted $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$, is:

$$MKLD(f,g) = \frac{1}{2}(\mathrm{tr}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - m + \ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|}) \quad (2)$$

where m is the dimension of the vector space, $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ are the vectors of the mean values.

3.2. Fault detection procedure

The general statistical monitoring procedure is based on the collection of a large number of healthy data samples used as the reference data set. All new measured data are then compared to the healthy ones to check whether an abnormal event (faulty situation) occurs. While the healthy data are not available, the problem can be considered as evaluating the difference between two consecutive moving temporal frames.

3.2.1. Univariate approach using KLD with PCA (PCA-KLD)

In the univariate approach, the healthy data are projected in the PCA framework, and a reference model of the PCA is obtained (eigenvectors). Then, a reference healthy probability distribution is estimated for each latent score. Afterwards, for each new set of observations, the associated latent scores are calculated through the predetermined PCA model and their probability distribution are estimated. Thus, the KLD is used to measure the dissimilarities between the probability density functions of healthy latent scores and measured ones. The obtained KLD value is compared finally to a threshold to make the decision (healthy or faulty). The choice of the threshold depends on the desired performances and the noise level in the system [12, 18].

3.2.2. Multivariate approach using MKLD for detection

For the multivariate approach, PCA is not used [19]. A reference multivariate probability distribution is estimated for the healthy data set. Then for each new measured set of observations, their multivariate probability distribution are computed, and compared to the reference ones by the MKLD. The MKLD value is then compared to a threshold to decide if the measured data set is healthy or faulty.

3.2.3. Comparison of performances

To compare the detection performances of the PCA-KLD with the MKLD, we consider as an example a multivariate Auto-Regressive (AR) system :

$$s(i) = \begin{bmatrix} 0.11 & -0.19 & 0.2 \\ 0.84 & 0.26 & 0.1 \\ 0.47 & -0.14 & 0.8 \end{bmatrix} s(i-1) + \begin{bmatrix} 1 & 2 & 7 \\ 3 & -4 & 4 \\ 6 & -7 & 5 \end{bmatrix} u(i-1)$$

$$y(i) = s(i) + v(i)$$

where u is the correlated input,

$$u(i) = \begin{bmatrix} 0.8 & 0.2 & 0.1 \\ 0.4 & 0.5 & 0.7 \\ 0.9 & 0.4 & 0.3 \end{bmatrix} u(i-1) + \begin{bmatrix} 0.1 & 0.6 & 0.4 \\ 0.3 & 0.7 & 0.2 \\ 0.1 & 0.5 & 0.7 \end{bmatrix} w(i-1)$$

w is a vector of 3 inputs $w = [w_1 \quad w_2 \quad w_3]^T$, which are independent Gaussian signals with zero mean and unit variance. $u = [u_1 \quad u_2 \quad u_3]^T$ is the vector of measured inputs, and $y = [y_1 \quad y_2 \quad y_3]^T$ is the vector of outputs corrupted by uncorrelated data Gaussian errors $v = [v_1 \quad v_2 \quad v_3]^T$ with zero mean and variance σ_v^2. The vector of process variables will be formed with the measured inputs and outputs of the process at instant i, i.e $[y_1 \quad y_2 \quad y_3 \quad u_1 \quad u_2 \quad u_3]^T$. A data matrix $\tilde{X}$ of N measurements/row is formed with these variables. We apply a gain fault (variance change) on the variable u_1.

To compare the detection limit of the two methods, we plot in Fig.1 and Fig.2 the ROC curves (detection probability, P_D along with false alarm probability, P_{FA}) considering a Signal to Noise Ratio (SNR) equal to 40dB, and for different Fault to Noise Ratio (FNR). For the PCA-KLD method, the first principal components representing the maximum data information are considered.

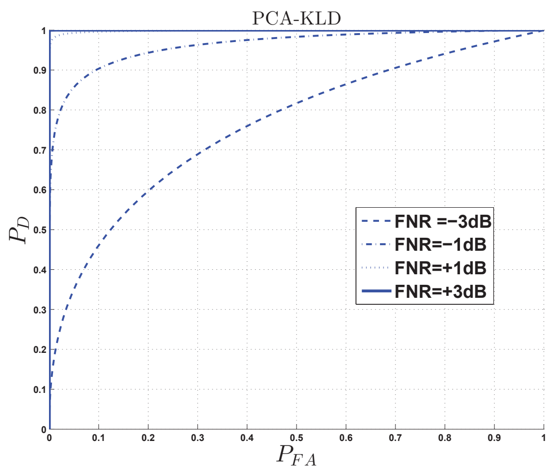

Fig. 1. PCA-KLD performance, SNR = 40dB

As seen in Fig.1, the PCA-KLD approach is efficient (no false alarm and 100 % detection) for FNR= +3dB and higher values. For lower FNR, performances are degraded. However, for the MKLD, Fig.2 shows that the ideal detection performance is attained from

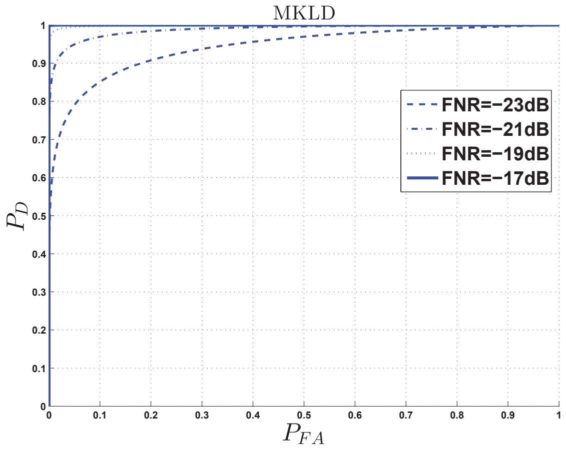

Fig. 2. MKLD performance, SNR = 40dB

FNR=-17dB (fault severity lower than the ideal PCA-KLD case (FNR=+3dB)).

In order to show the noise effect on both methods, we consider a gain fault of severity $a = 0.002$ for several noise conditions (SNR = [20dB, 25dB, 30dB, 40dB]). The results are given in Fig.3. As seen in figure 3, the fault can be easily detected by the MKLD

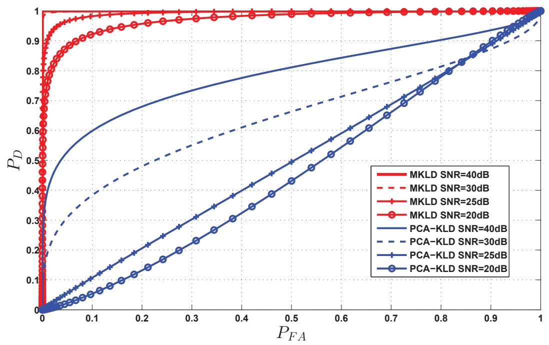

Fig. 3. MKLD vs PCA-KLD detection performance comparison

for SNR=40dB and SNR=30dB. For SNR=20dB and SNR=25dB the performances of the MKLD is affected and the detection are lowed but still good. However, for the PCA-KLD approach the detection performances are more severely affected by the noise for the considered SNR. The more the noise variance is high, lower are the detection performances.
Therefore, the MKLD widely improves the detection performance compared to the univariate approach and this detection is also sufficiently robust to noise influence.

4. FAULT ESTIMATION

Once a fault occurrence is detected, its severity (amplitude) should be estimated for diagnosis purpose. As shown in the previous section, the MKLD values obtained for two different fault amplitudes, in the same noise conditions, are not the same, therefore the MKLD can be used for fault estimation. In order to estimate the fault amplitude (severity), a model of the MKLD function of the fault amplitude and the noise level has to be determined. Consequently, in this section, we theoretically compute the MKLD model in terms of the fault amplitude denoted a and the original data set parameters ($MKLD = f(a, \theta_{\tilde{X}}, \theta_X)$) where $\theta_{\tilde{X}}$ and θ_X are the probability distribution parameters of the healthy and the faulty data set respectively. Thus, the fault amplitude is estimated using the inverse model ($\hat{a} = f^{-1}(MKLD, \theta_{\tilde{X}}, \theta_X)$).

4.1. MKLD modeling for incipient fault

Considering that the healthy data set $\tilde{X}$ follows a Gaussian distribution $f(\tilde{X}) \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, and the faulty data set X follows a Gaussian distribution $g(X) \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$.
$\boldsymbol{\Sigma}_0$ is the covariance matrix of the healthy data:

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m1} & \dots & \dots & c_{mm} \end{pmatrix} \tag{3}$$

$\boldsymbol{\mu}_0$ is the mean vector of the original data:

$$\boldsymbol{\mu}_0 = (\mu_{01} \quad \mu_{02} \quad .. \quad \mu_{0m})^T \tag{4}$$

We assume that one variable x_j is affected by a gain fault such as:

$$x_j = G \times \tilde{x}_j \tag{5}$$

where $\tilde{x}_j$ is the reference variable and $G = 1 + a$.
The covariance matrix $\boldsymbol{\Sigma}_1$ of the faulty data set is:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} c_{11} & c_{12} & \dots & Gc_{1j} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & Gc_{2j} & \dots & c_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Gc_{j1} & Gc_{j2} & \dots & G^2 c_{jj} & \dots & Gc_{jm} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{m1} & \dots & \dots & Gc_{mj} & \dots & c_{mm} \end{pmatrix} \tag{6}$$

Let us define the matrix $A_{[m \times m]}$ as:

$$A = \begin{pmatrix} 1 & 0 & \dots & 0^{\downarrow j} & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & G & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{pmatrix} \tag{7}$$

We can write:

$$\boldsymbol{\Sigma}_1 = A \times \boldsymbol{\Sigma}_0 \times A \tag{8}$$

and

$$\boldsymbol{\Sigma}_1^{-1} = A^{-1} \times \boldsymbol{\Sigma}_0^{-1} \times A^{-1} \tag{9}$$

$\boldsymbol{\Sigma}_0^{-1}$ is the inverse of the reference covariance matrix :

$$\boldsymbol{\Sigma}_0^{-1} = \begin{pmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_{m1} & \dots & \dots & \delta_{mm} \end{pmatrix} \tag{10}$$

$$\boldsymbol{\Sigma}_1^{-1} \times \boldsymbol{\Sigma}_0 = A^{-1} \times \boldsymbol{\Sigma}_0^{-1} \times A^{-1} \times \boldsymbol{\Sigma}_0 \tag{11}$$

Then

$$tr(\boldsymbol{\Sigma}_1^{-1} \times \boldsymbol{\Sigma}_0) = tr(A^{-1} \times \boldsymbol{\Sigma}_0^{-1} \times A^{-1} \times \boldsymbol{\Sigma}_0) \tag{12}$$

As the trace function is the sum of the diagonal elements in a matrix, we should then multiply the 2 matrices $(A^{-1} \times \mathbf{\Sigma}_0^{-1})$ and $(A^{-1} \times \mathbf{\Sigma}_0)$ to obtain the diagonal elements. We finally, obtain:

$$tr(\mathbf{\Sigma}_1^{-1} \times \mathbf{\Sigma}_0) = 2\sum_{i=1}^{m-1}\sum_{n=i+1}^{m} \delta_{in}c_{in} + \sum_{i=1}^{m}\delta_{ii}c_{ii} + \frac{1}{G^2}\delta_{jj}c_{jj} + \frac{2}{G}\sum_{i=1}^{m}\delta_{ij}c_{ij} \quad (i \ \& \ n \neq j) \quad (13)$$

The faulty mean vector μ_1 can be written after a gain fault as:

$$\boldsymbol{\mu}_1 = (\mu_{01} \quad \mu_{02} \quad .. \quad G\mu_{0j} \quad .. \quad \mu_{0m})^T \quad (14)$$

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = (0 \quad 0 \quad .. \quad (G-1)\mu_{0j} \quad .. \quad 0)^T \quad (15)$$

Then:

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^{\mathrm{T}} \mathbf{\Sigma}_1^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = (\frac{G-1}{G}\mu_{0j})^2 \times \delta_{jj} \quad (16)$$

We must now compute the function $\ln \frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|}$:

$$|\mathbf{\Sigma}_1| = |A \times \mathbf{\Sigma}_0 \times A| = |A| \times |\mathbf{\Sigma}_0| \times |A| \quad (17)$$

$$\ln \frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|} = \ln \frac{|A| \times |\mathbf{\Sigma}_0| \times |A|}{|\mathbf{\Sigma}_0|} = \ln |A|^2 = \ln G^2 \quad (18)$$

Therefore, based on (13), (16) and (18), the analytical model of the MKLD for fault diagnosis purpose can be written as (19).

4.2. Model validation

To validate the model, we use the same numerical example already described in section 3.2.3. The fault affects the variable u_1, *i.e* the fourth variable of the healthy database $\tilde{X}$.

In Fig.4, MKLD is computed along with fault severity for different SNR using the direct expression (direct model (2)) and the analytical model (fault diagnosis model (19)). As seen in this figure, the theoretical model for diagnosis perfectly fits with the direct expression while the noise level is low or null. Nevertheless, the more the noise level is important, higher will be the difference between the direct model and the MKLD analytical model for diagnosis.

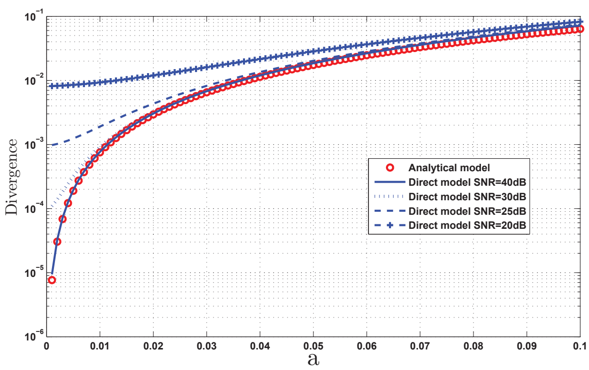

Fig. 4. Analytical Model Validation

4.3. Fault estimation: Inverse model

By inverting the analytical model (19), the theoretical estimation of the fault amplitude that depends on the divergence value $M\hat{K}LD$ is finally given by (20), where :

$$\beta = 2\sum_{i=1}^{m-1}\sum_{n=i+1}^{m} \delta_{in}c_{in} + \sum_{i=1}^{m}\delta_{ii}c_{ii} + \mu_{0j}^2\delta_{jj} - m - 2 - 2 \times M\hat{K}LD$$

$$\gamma = 2\sum_{i=1}^{m}\delta_{ij}c_{ij} - 2\mu_{0j}^2 \times \delta_{jj} \quad and \quad \eta = c_{jj}\delta_{jj} + \mu_{0j}^2\delta_{jj}$$

To validate the estimation, we apply a gain fault on u_1, with an amplitude a in the range $a = [0, .., 0.1]$. In Fig.5, we plot the estimated fault amplitude $\hat{a}$ for different SNR values.

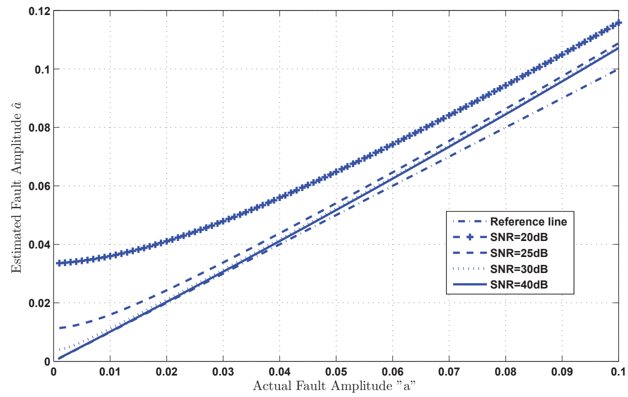

Fig. 5. Fault amplitude estimation results

We can first notice on this figure that we have for all the SNR values an overestimation of the actual fault amplitude. This overestimation provides a safety margin in a sensible fault diagnosis context. Secondly, it must be noticed that the estimated fault amplitude is close to the actual one for SNR=30dB and SNR=25dB. However when the SNR decreases (SNR<25dB), the error increases and becomes important especially for small faults ($a < 0.02$).

To study the accuracy of the estimation we have computed the estimate's relative error. The estimation of the faulty variable u_1, *i.e* x_4 is $\hat{x}_4 = (1 + \hat{a}) \times (x_4)$. The relative error, denoted ϵ_a is: $\epsilon_a = \frac{\hat{x}_4 - x_4}{x_4} = \frac{\hat{a} - a}{1 + a}$.

In Fig.6, the estimate's relative error is plotted along with the fault severity for different SNR values. We can see that the relative error is very small even for SNR=20dB and for small fault amplitude. We can notice that the variation of ϵ_a for all the considered SNR values is always lower than 3.5%. For small fault amplitude, ϵ_a decreases along with a for $a \in [0, 0.02]$. Afterwards, ϵ_a slightly increases for $a > 0.02$. However, this modeling error is still lower than 1.5%. The evolution of the estimation error is mainly due to the small slope value in the MKLD function for high fault amplitude, so a very small modeling error increases the estimation error.

5. CONCLUSION

In this paper we have proposed a fault detection and estimation approach using MKLD. The method allows, for any distributions, the detection of incipient faults more accurately than other approaches

$$MKLD(f,g) = \frac{1}{2}\left\{2\sum_{i=1}^{m-1}\sum_{n=i+1}^{m}\delta_{in}c_{in} + \sum_{i=1}^{m}\delta_{ii}c_{ii} + \frac{1}{G^2}\delta_{jj}c_{jj} + \frac{2}{G}\sum_{i=1}^{m}\delta_{ij}c_{ij} + \left(\frac{G-1}{G}\mu_{0j}\right)^2 \times \delta_{jj} - m + \ln G^2\right\} \quad (i \;\&\; n \neq j) \tag{19}$$

$$\hat{a} = \frac{\sqrt[3]{\sqrt{(-108\eta + 18\beta\gamma - 2\beta^3)^2 + 4(6\gamma-\beta)^3} - 108\eta^2 + 18\beta\gamma - 2\beta^3}}{6\sqrt[3]{2}} \tag{20}$$

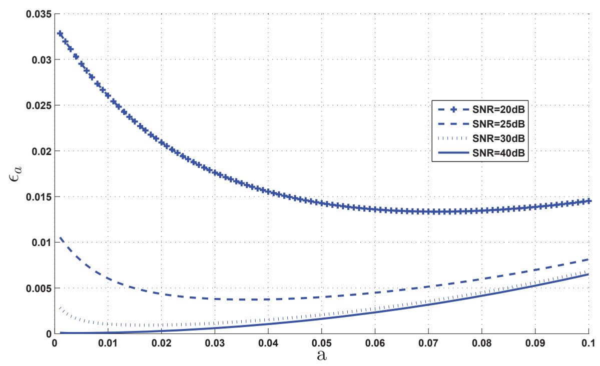

Fig. 6. Relative estimation error

based on PCA and fault with lower severity can be considered. For estimating the fault severity, we have developed an analytical model of the MKLD (in the Gaussian case) from which the fault amplitude expression is derived. The model has been validated for different noise conditions and its accuracy has been evaluated . The results prove that for incipient faults, severity estimation based on this model is accurate with a low relative error and a safety margin thanks to the overestimation of the actual fault amplitude. For future works, the sensitivity of our approach to the used estimators will be studied.

6. REFERENCES

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes-Theory and Application*, Prentice Hall information and system sciences series. Prentice Hall,Englewood Cliffs, NJ, 1993.

[2] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Advanced Textbooks in Control and Signal Processing. London: Springer, 2001.

[3] J. F. MacGregor and T. Kourti, “Statistical process control of multivariate processes,” *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995.

[4] V. Venkatasubramanian, R. Raghunathan, N. K. Surya, and Y. Kewen, “A review of process fault detection and diagnosis: Part III: Process history based methods,” *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003.

[5] M. Harkat, G. Mourot, and J. Ragot, “An improved PCA scheme for sensor FDI: application to an air quality monitoring network,” *J. of Process Control*, vol. 16, pp. 625–634, 2006.

[6] I.T. Jolliffe, *Principal Component Analysis (Second Edition)*, Springer Series in Statistics. London: Springer, 2002.

[7] U. Kruger and L. Xie, *Advances in Statistical Monitoring of Complex Multivariate Processes*, New York : Wiley, 2012.

[8] S. Wang and Y. Chen, “Sensor validation and reconstruction for building central chilling systems based on principal component analysis,” *Energy Conversion and Management*, vol. 45, no. 5, pp. 673–695, March 2004.

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

[10] J. Harmouche, C. Delpha, and D. Diallo, “Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part I,” *Signal Processing, Elsevier*, vol. 94, pp. 278–287, January 2014.

[11] J. Harmouche, C. Delpha, and D. Diallo, “Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part II,” *Signal Processing, Elsevier*, vol. 109, pp. 334–344, April 2015.

[12] A. Youssef, C. Delpha, and D. Diallo, “An optimal fault detection threshold for early detection using Kullback-Leibler divergence for unknown distribution data,” *Signal Processing, Elsevier*, vol. 120, pp. 266–279, March 2016.

[13] H. Ringberg, A. Soule, J. Rexford, and C. Diot, “Sensitivity of PCA for traffic anomaly detection,” in *Proc. of the 2007 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, New York, USA, June 2007.

[14] B. Wang, Q. Jiang, and X. Yan, “Fault detection and identification using a Kullback-Leibler divergence based multi-block principal component analysis and bayesian inference,” *Korean Journal of Chemical Engineering*, vol. 31, no. 6, pp. 930–943, June 2014.

[15] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, pp. 79–86, 1951.

[16] M.T. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2nd edition, USA, 2005.

[17] J.E. Contreras-Reyes and R.B. Arellano-Valle, “Kullback-Leibler divergence measure for multivariate skew-normal distributions,” *Entropy*, vol. 14, pp. 1606–1626, 2012.

[18] 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 Industrial Electronics Society Conference (IECON 2013)*, Vienna, Austria, November 10-13 2013, pp. 7364 – 7369.

[19] J. Zeng, U. Kruger, J. Geluk, X. Wang, and L. Xie, “Detecting abnormal situations using the Kullback-Leibler divergence,” *Automatica*, vol. 50, no. 11, pp. 2777–2786, November 2014.