ANALYTICAL MODEL OF THE KL DIVERGENCE FOR GAMMA DISTRIBUTED DATA: APPLICATION TO FAULT ESTIMATION

Abdulrahman Youssef ⋆,⋄ *Claude Delpha* ⋆ *Demba Diallo* ⋄

⋆ Laboratoire des Signaux et Systemes (L2S), Univ. Paris-Sud, CNRS, CentraleSupelec
⋄ Group of Electrical Engineering Paris (GEEPS), CNRS, CentraleSupelec, Univ. Paris-Sud, UPMC
91192 Gif Sur Yvette, France

ABSTRACT

Incipient fault diagnosis has become a key issue for reliability and safety of industrial processes. Data-driven methods are effective for feature extraction and feature analysis using multivariate statistical techniques. Beside fault detection, fault estimation is essential for making the appropriate decision (safe stop or fault accommodation). Therefore, in this paper, we have developed an analytical model of the Kullback-Leibler Divergence (KLD) for Gamma distributed data to be used for the fault severity estimation. In the Principal Component Analysis (PCA) framework, the proposed model of the KLD has been analysed and compared to an estimated value of the KLD using the Monte-Carlo estimator. The results show that for incipient faults ($<10\%$) in usual noise conditions (SNR$>$40dB), the analytical model is accurate enough with a relative error around 10%.

***Index Terms*—** Fault detection, KLD model and estimation, Gamma distributed data, Incipient faults.

1. INTRODUCTION

Increased productivity requirements and stringent performance specifications lead to more demanding operating conditions of many modern industrial systems. Such conditions increase the possibility of system failures. Sensor, actuator or plant failures may drastically change the system behavior, resulting in degradation or even instability. In order to improve the efficiency, the reliability can be achieved by fault tolerant control (FTC) which includes two main approaches : passive FTC and active FTC [1].

In passive FTC, the robust controller is built to cope with the healthy mode and also accommodate the faults. However, all the possible faults scenarios must be known in advance and the input evaluated, which is very conservative.

Active FTC includes a Fault Detecion and Isolation (FDI) module and supervisor, which decides to maintain the same controller with adequate parameters (reconfiguration) or engage a new controller (restructuring) to guarantee a required level of reliability and safety [1].

FDI has become an attractive topic and has received considerable attention during the past two decades because the efficacy of the fault management relies on the accuracy of the fault detection and the fault severity assessment. Indeed fault detection and fault estimation methods should be robust to noise and unexpected uncertainties and perturbations. If an accurate analytical fault model is available, optimization-based approach can be used to estimate the fault amplitude. But in most case, a data-driven model is used combined with statistical related methods.

In this paper, we adopt a data-driven approach using descriptive features within a Statistical Process Control (SPC) usual technique, the Principle Component Analysis (PCA) [2], [3], framework combined with multivariate statistical techniques to develop an efficient fault detection and estimation method. 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 [2]. In addition, these process monitoring techniques are attractive for industrial practical processes because they only require a good historical data set of healthy operation, which are easily obtainable for computer-controlled industrial processes. PCA-based monitoring methods and their extensions have been successfully applied in a wide range of applications and industries, such as in chemical processes, air and water treatment, transport systems, energy, medical devices, and many others [4].

It has already been shown, that in the PCA framework, the Kullback-Leibler Divergence (KLD) [5] is conceptually more straightforward and also more sensitive for the fault detection of incipient faults [6], [7] than the usual detection indices, like the Hotelling T^2 and squared prediction error (SPE).

The goal of this work is then to derive from data an analytical model of the KLD that will be used for incipient fault estimation. We assume in the following that the data are multivariate gamma distributed. Indeed, this assumption is not too restrictive as this distribution encompasses the Gaussian and the χ^2 ones which can be found in many areas as for example acoustics vibration processing. For this work, we compare the KLD estimation using Monte Carlo simulation and the KLD model based on Gamma distribution to show the efficiency of the proposed model including the noise influence.

2. DEFINITION

2.1. Notation

Let's introduce the following notations:
We can set $X_{[N \times m]}$ such as $X = (x_1, ..., x_i, ..., x_m)$ is the original data matrix where $x_i = [x_{1i} ... x_{Ni}]'$ is a column vector of N measurements taken for the ith variable.
$\bar{X}_{[N \times m]}$, for $\bar{X} = (\bar{x}_1, ..., \bar{x}_i, ..., \bar{x}_m)$ is the centered matrix.
$S_{[m \times m]} = \frac{1}{N} \bar{X}' \bar{X}$ 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 $T = \bar{X} P$.
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 for the data samples.

2.2. Signals Model

While considering a random variable u that is Gamma-distributed with shape α and scale β denoted by:

$$u \sim \Gamma(\alpha, \beta) \tag{1}$$

The probability density function using the shape-scale parametrization is:

$$f(u; \alpha, \beta) = \frac{u^{\alpha - 1} e^{-\frac{u}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)} \quad \text{for } u > 0 \text{ and } \alpha, \beta > 0 \tag{2}$$

Here $\Gamma(\alpha)$ is the Gamma function evaluated at α.
Considering $v_{[N \times m]}$ assumed to be a white Gaussian noise with a normal distribution ($\sim \mathcal{N}(0, \sigma_v^2)$), every column of v is the additive noise of the corresponding signal in X.
In our work, we suppose that the noise model corresponds to an environmental change between healthy data measurements and the test measured data. Then, the noise is added only on the test data and the healthy data are noise free. Moreover, all the noises components on the m signals variables are assumed to be the same (columns of the matrix v are supposed equal). This assumption is possible as we consider that the m signals are measured at the same time.
We can then write in case of fault free test data (healthy):

$$x_i = x_i^* + v_i \tag{3}$$

As said before, our signal components x_i^* follows a Gamma distribution $x_i^* \sim \Gamma(\alpha_i^*, \beta_i^*)$. We consider that the noisy signal will also follow a Gamma distribution. Then $x_i \sim \Gamma(\alpha_i, \beta_i)$. We can calculate α_i and β_i with respect to α_i^*,β_i^* and σ_v^2 by identification of the first and the second moment.

$$\alpha_i = \frac{(\alpha_i^* \beta_i^*)^2}{\alpha_i^* (\beta_i^*)^2 + \sigma_v^2} \tag{4}$$

$$\beta_i = \frac{\alpha_i^* (\beta_i^*)^2 + \sigma_v^2}{\alpha_i^* \beta_i^*} \tag{5}$$

2.3. Incipient Fault Model

Incipient faults are defined as slowly developing faults or slight unpredictable changes in the system. They are characterized by a small amplitude compared to the useful signal [8]. In a short time window, the incipient fault amplitude is assumed to be a constant a.
Hereafter, we consider a gain fault. Faulty components are therefore proportional to the reference signal ones.
Let's assume that the considered fault occurs only on one descriptive feature (variable) among the m measured ones. The fault affecting the jth variable x_j among the m process variables can be written as:

$$f_a = a \times \left[\begin{bmatrix} x_{1j}^* \\ \vdots \\ \vdots \\ x_{Nj}^* \end{bmatrix} + \begin{bmatrix} v_{1j} \\ \vdots \\ \vdots \\ v_{Nj} \end{bmatrix} \right] \tag{6}$$

Where f_a is the fault component, a is the fault amplitude parameter and x_j^* is the jth reference signal which will be affected by the fault.
With such signal fault and noise modelling, we propose to study the Kullback-Leibler divergence for the fault detection.

2.4. Kullback-Leibler Divergence for Detection

The difference between the probability density functions (PDFs) of healthy and test data can be achieved by the KLD computation between the two distributions [5].
For discrimination between two continuous probability density functions (PDFs) $f(u)$ and $g(u)$ of a random variable u, the Kullback-Leibler Information (KLI) is defined as:

$$I(f \| g) = \int f(u) \log \frac{f(u)}{g(u)} du. \tag{7}$$

The KL Divergence (KLD) is then defined as the symmetric version of the KL Information [5], [9]:

$$KLD(f, g) = I(f \| g) + I(g \| f). \tag{8}$$

KLD approximation by Monte Carlo Simulation
For arbitrary distributions f and g, (7) can be numerically approximated using Monte Carlo (MC) simulation. The Monte Carlo method expresses (7) 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:

$$K\widehat{L}D(f, g) = D_{MC}(f, g) = \frac{1}{n_s} \sum_{i=1}^{n_s} log \frac{f(z_i)}{g(z_i)} \tag{9}$$

With such approximation, the estimation error distribution is normal with variance σ_{MC}^2 and zero mean ($\sim \mathcal{N}(0, \sigma_{MC}^2)$) such as $\sigma_{MC}^2 = \frac{1}{n_s} \mathbb{E}([log(f/g)]^2)$. More the set of samples n_s will be larger, smaller will be the Monte Carlo estimation error.

3. FAULT DETECTION AND ESTIMATION PROCEDURE

The general procedure of statistical monitoring is to collect 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 behavior occurs. 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 and their probability distribution are estimated. Then, the KLD is used to measure the dissimilarities between the probability density functions of healthy latent scores and measured ones. In a complex system, where the number of the representative latent scores is high, its is not practical to apply the KLD for each latent score. In this case, a multivariate KLD should be applied to measure the dissimilarities between the latent scores together. The main steps of this fault detection and estimation procedure are shown in the flow-chart depicted in Fig.1.

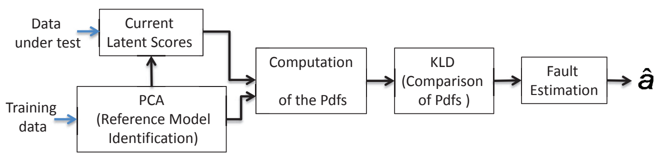

Fig. 1. KLD-based FDI flowchart

3.1. KLD modelling

For Gamma densities f and g such that f $\sim \Gamma(\alpha_1, \beta_1)$ and $g \sim \Gamma(\alpha_2, \beta_2)$ the Kullback-Leibler Divergence between f and g is given by:

$$KLD(\alpha_1, \beta_1; \alpha_2, \beta_2) = (\alpha_1 - \alpha_2)\psi(\alpha_1) - \log\Gamma(\alpha_1) \quad (10)$$

$$+\log\Gamma(\alpha_2) + \alpha_2(\log\beta_2 - \log\beta_1) + \alpha_1\frac{\beta_1 - \beta_2}{\beta_2}$$

where $\psi(\alpha)$ is the digamma function.
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 Gamma distributed. So principal component scores, which are linear combinations of the original variables, can be approximated by Gamma distributed variables with a negligible error [10].
From the assumption of Gamma distribution, it follows that each of the l principal scores t_k, $(k = 1, 2, ..., l)$, has a PDF which we denote f such that $f \sim \Gamma(\alpha_k, \beta_k)$. We propose to compare f against its reference. The reference is denoted f^* such as $f^* \sim \Gamma(\alpha_k^*, \beta_k^*)$.

3.1.1. Computation of the reference PDF parameters α_k^*, β_k^*

$$t_k^* = X^*_{[N\times m]} \times p_{k[m\times 1]} \quad (11)$$

where p_k is the k_{th} eigenvector of the data matrix X^*.

$$p_{k[m\times 1]} = [p_{1k} \quad p_{2k} \quad \dots \quad p_{ik} \quad \dots \quad p_{mk}]' \quad (12)$$

$$t_k^* = p_{1k}x_1^* + p_{2k}x_2^* + \dots + p_{ik}x_i^* + \dots + p_{mk}x_m^* \quad (13)$$

We have:

$$x_i^* \sim \Gamma(\alpha_i^*, \beta_i^*)$$

$$p_{ik}x_i^* \sim \Gamma(\alpha_i^*, p_{ik} \times \beta_i^*)$$

$$t_k^* = \sum_{i=1}^{i=m} p_{ik}x_i^* \sim \Gamma(\alpha_k^*, \beta_k^*)$$

Here we have a sum of Gamma random variables. We can approximate the sum by a Gamma distribution $\Gamma(\alpha_k^*, \beta_k^*)$. The shape α_k^* and the scale β_k^* are obtained by identification of the 2 first statistical moments (mean and variance).

$$\alpha_k^* = \frac{\mu^2_{t_k^*}}{\sigma^2_{t_k^*}} \qquad \beta_k^* = \frac{\sigma^2_{t_k^*}}{\mu_{t_k^*}} \quad (14)$$

The problem is now to estimate the mean and the variance of the principle component t_k^*:

$$\mu_{t_k^*} = \mathbb{E}(\sum_{i=1}^{i=m} p_{ik}x_i^*) = \sum_{i=1}^{i=m}\mathbb{E}(p_{ik}x_i^*) = \sum_{i=1}^{i=m} p_{ik}\alpha_i^*\beta_i^* \quad (15)$$

$$\sigma^2_{t_k^*} = \mathbb{E}((\sum_{i=1}^{i=m} p_{ik}x_i^*) - \mu_{t_k^*})^2 \quad (16)$$

Using the covariance properties, we can obtain the expression of $\sigma^2_{t_k^*}$:

$$\sigma^2_{t_k^*} = \sum_{i=1}^{i=m}\alpha_i^* p_{ik}^2(\beta_i^*)^2 + 2\sum_{i=1}^{i=m-1}\sum_{q=i+1}^{q=m} p_{ik}p_{qk}S_{iq} \quad (17)$$

Finally, introducing (15) and (17) into (14) we obtain:

$$\alpha_k^* = \frac{(\sum_{i=1}^{i=m}\alpha_i^* p_{ik}\beta_i^*)^2}{\sum_{i=1}^{i=m}\alpha_i^* p_{ik}^2(\beta_i^*)^2 + 2\sum_{i=1}^{i=m-1}\sum_{q=i+1}^{q=m} p_{ik}p_{qk}S_{iq}} \quad (18)$$

$$\beta_k^* = \frac{\sum_{i=1}^{i=m}\alpha_i^* p_{ik}^2(\beta_i^*)^2 + 2\sum_{i=1}^{i=m-1}\sum_{q=i+1}^{q=m} p_{ik}p_{qk}S_{iq}}{\sum_{i=1}^{i=m}\alpha_i^* p_{ik}\beta_i^*} \quad (19)$$

Where

$$\alpha_i^* = \frac{\mu^2_{x_i^*}}{\sigma^2_{x_i^*}} = \frac{(\frac{1}{N}\sum_{r=1}^{N} x_{ri}^*)^2}{\frac{1}{N}\sum_{r=1}^{N}(x_{ri}^* - \frac{1}{N}\sum_{r=1}^{N} x_{ri}^*)^2} \quad (20)$$

$$\beta_i^* = \frac{\sigma^2_{x_i^*}}{\mu_{x_i^*}} = \frac{\frac{1}{N}\sum_{r=1}^{N}(x_{ri}^* - \frac{1}{N}\sum_{r=1}^{N} x_{ri}^*)^2}{\frac{1}{N}\sum_{r=1}^{N} x_{ri}^*} \quad (21)$$

3.1.2. *Computation of the faulty PDF parameters* α_k, β_k

As mentioned before, the fault affects one variable x_j. We can write:

$$x_j = (x_j^* + v_j) + a \times (x_j^* + v_j) \tag{22}$$

where a is the fault amplitude.
Therefore, if the noisy variable $(x_j^* + v_j) \sim \Gamma(\alpha_j, \beta_j)$ then the resulting faulty variable is $x_j \sim \Gamma(\alpha_j, (1+a) \times \beta_j)$
In the PCA framework the scores can be written as:

$$t_k = p_{1k}x_1 + p_{2k}x_2 + ... + p_{jk}x_j + ... + p_{mk}x_m \quad \sim \Gamma(\alpha_k, \beta_k) \tag{23}$$

Then, based on the Gamma approximation of a sum of Gamma random variables, the expression of α_k and β_k are given in (24) and (25).

$$\alpha_i = \frac{\mu_{x_i}^2}{\sigma_{x_i}^2} = \frac{(\frac{1}{N}\sum_{r=1}^{N} x_{ri}^*)^2}{\frac{1}{N}\sum_{r=1}^{N}(x_{ri}^* - \frac{1}{N}\sum_{r=1}^{N} x_{ri}^*)^2 + \sigma_v^2} \tag{26}$$

$$\beta_i = \frac{\sigma_{x_i}^2}{\mu_{x_i}} = \frac{\frac{1}{N}\sum_{r=1}^{N}(x_{ri}^* - \frac{1}{N}\sum_{r=1}^{N} x_{ri}^*)^2 + \sigma_v^2}{\frac{1}{N}\sum_{r=1}^{N} x_{ri}^*} \tag{27}$$

Let's define:

$$\delta_y = \sum_{i=1}^{i=m} \alpha_i p_{ik} \beta_i \qquad \delta_y^* = \sum_{i=1}^{i=m} \alpha_i^* p_{ik} \beta_i^* \tag{28}$$

$$\delta_w = \sum_{i=1}^{i=m} \alpha_i p_{ik}^2 (\beta_i)^2 \qquad \delta_w^* = \sum_{i=1}^{i=m} \alpha_i^* p_{ik}^2 (\beta_i^*)^2$$

$$\theta_y = \alpha_j p_{jk} \beta_j \qquad \theta_w = \alpha_j p_{jk}^2 (\beta_j)^2$$

Then the KLD model becomes as (29).
Inverting the KLD model in (29) the theoretical estimation of the fault amplitude that depends on the divergence value is finally given by:

$$\widehat{a} = \frac{\theta_y - 2\theta_w + \sqrt{(2\theta_w - \theta_y)^2 - 4(\theta_w \delta_w - \xi \delta_y)}}{2\theta_w} \tag{30}$$

where

$$\xi = \frac{\beta_k^*}{e^{-\sqrt{\frac{2\widehat{KLD}}{\alpha_k^*}}}} \tag{31}$$

3.2. Model validation

We consider here a multivariate AR system, in which the generated signals are approximately Gamma distributed. It is defined at instant h as follows:

$$c(h) = \begin{pmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{pmatrix} c(h-1) + \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix} q(h-1) \tag{32}$$

$$y(h) = c(h) + \rho(h)$$

where q is the correlated input,

$$q(h) = \begin{pmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{pmatrix} q(h-1) + \begin{pmatrix} 0.193 & -0.689 \\ -0.32 & -0.749 \end{pmatrix} d(h-1) \tag{33}$$

d is a vector of 2 inputs $d = [d_1 \quad d_2]'$, which are uncorrelated Gamma signals ($\sim \Gamma(0.5, 0.5)$). $q = [q_1 \quad q_2]'$ is the vector of measured inputs, and $y = [y_1 \quad y_2]'$ is the vector of outputs corrupted by uncorrelated Gamma error $\rho = [\rho_1 \quad \rho_2]'$ ($\sim \Gamma(0.1, 0.2)$). The vector of process variables will be formed with the measured inputs and outputs of the process at instant h, i.e $[y_1(h) \quad y_2(h) \quad q_1(h) \quad q_2(h)]$. A data matrix X of N measurements/rows (N=10^6) is formed with these variables. In our example, we will consider that the fault affects only the variable q_2. PCA is applied on the data covariance matrix; 4 principal components are obtained with variance $\lambda = [40.26 \quad 4.9 \quad 1.14 \quad 0.17]$. The first principal component accounts for 86.6% of variance.
For our study, we consider the first component t_1^* where the corresponding eigenvalue is the highest one, we obtain $t_1^* = Xp_1^*$. Then, the probability density of t_1^* is estimated as the reference distribution and the probability density of the faulty t_1 is estimated as the test distribution. Fig.2 displays the KLD computed on the first principle component t_1 versus the fault amplitude a for different noise levels. As seen in this figure, the theoretical model fits very well with the approximated divergence.

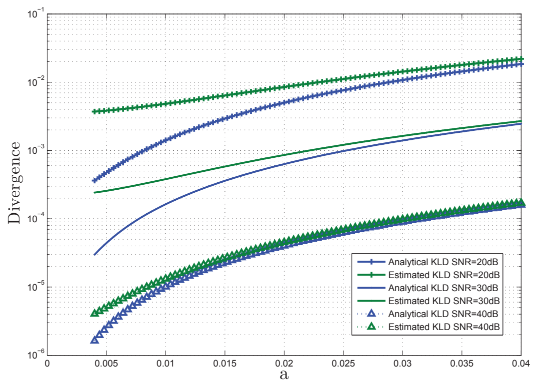

Fig. 2. Comparison of KLD and $\widehat{KLD}$

Nevertheless, the more the noise power is important, higher will be the difference between the estimated KLD and the analytical model.

Fig.3 shows the relative error (ϵ_{KLD}) between the analytical model and the estimated KLD for different noises and number of samples conditions. This relative error can be denoted $\epsilon_{KLD} = \frac{\widehat{KLD} - KLD}{\widehat{KLD}}$.

$$\alpha_k = \frac{((\sum_{i=1}^{i=j-1} \alpha_i p_{ik}\beta_i) + \alpha_j p_{jk}(1+a)\beta_j + (\sum_{i=j+1}^{i=m} \alpha_i p_{ik}\beta_i))^2}{(\sum_{i=1}^{i=j-1} \alpha_i p_{ik}^2 \beta_i^2) + \alpha_j p_{jk}^2 (1+a)^2 (\beta_j)^2 + (\sum_{i=j+1}^{i=m} \alpha_i p_{ik}^2 \beta_i^2) + 2\sum_{i=1}^{i=m-1}\sum_{q=i+1}^{q=m} p_{ik} p_{qk} S_{iq}} \quad (24)$$

$$\beta_k = \frac{(\sum_{i=1}^{i=j-1} \alpha_i p_{ik}^2 \beta_i^2) + \alpha_j p_{jk}^2 (1+a)^2 (\beta_j)^2 + (\sum_{i=j+1}^{i=m} \alpha_i p_{ik}^2 \beta_i^2) + 2\sum_{i=1}^{i=m-1}\sum_{q=i+1}^{q=m} p_{ik} p_{qk} S_{iq}}{(\sum_{i=1}^{i=j-1} \alpha_i p_{ik}\beta_i) + \alpha_j p_{jk}(1+a)\beta_j + (\sum_{i=j+1}^{i=m} \alpha_i p_{ik}\beta_i)} \quad (25)$$

$$KLD = (\frac{\delta_y^{*2}}{\delta_w^*} - \frac{(\delta_y+\theta_y a)^2}{\delta_w + \theta_w a^2 + \theta_w 2a})\psi(\frac{\delta_y^{*2}}{\delta_w^*}) - \log\Gamma(\frac{\delta_y^{*2}}{\delta_w^*}) + \log\Gamma(\frac{(\delta_y+\theta_y a)^2}{\delta_w + \theta_w a^2 + \theta_w 2a}) + \quad (29)$$
$$\frac{(\delta_y+\theta_y a)^2}{\delta_w + \theta_w a^2 + \theta_w 2a}(\log(\frac{\delta_w + \theta_w a^2 + \theta_w 2a}{\delta_y + \theta_y a}) - log(\frac{\delta_w^*}{\delta_y^*})) + \frac{\delta_y^{*2}}{\delta_w^*}(\frac{\delta_w^*(\delta_y+\theta_y a)}{\delta_y^*(\delta_w + \theta_w a^2 + 2\theta_w a)} - 1)$$

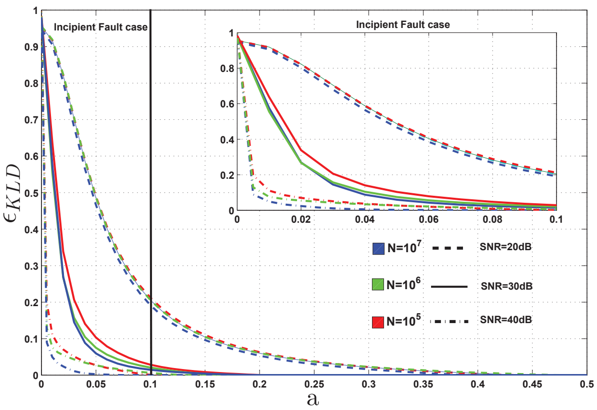

Fig. 3. KLD model relative error compared to estimation

Fig.3 is plotted for 3 noise levels (SNR = [$20dB$, $30dB$, $40dB$]) and 3 number of samples (N = [$10^5, 10^6, 10^7$]). We can draw the following remarks :

- In severe noise conditions, (SNR≤20dB) even with an important number of samples, the relative error is high for incipient faults ($a < 0.1$)
- In usual conditions (SNR≥40dB) the analytical model is accurate enough for the three number of samples.

4. CONCLUSION

Within the fault diagnosis process, fault estimation is crucial for making the appropriate decision. This task can be tedious particularly for incipient fault (small fault amplitude in noisy environment). For Gamma distributed data, we have developed in the PCA framework, an analytical model of the KLD from which the fault amplitude expression is derived. This model has been validated through a comparison with a numerically estimated KLD for different noise conditions and for different data size. The results show that for incipient faults, in usual noise conditions (SNR≥40dB), the model is accurate with a relative error lower than 10%.

REFERENCES

[1] Y. Diao and K. Passino, "Intelligent fault-tolerant control using adaptive and learning methods," *Control Engineering Practice*, vol. 10, pp. 801–817, 2002.

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

[3] A. Wiesel and A. O. Hero III, "Decomposable principal component analysis," *IEEE Trans. on Signal Processing*, vol. 57, no. 11, pp. 4369–4377, Nov. 2009.

[4] F. Harrou, M. N. Nounou, H. N. Nounou, and M. Madakyaru, "Statistical fault detection using PCA-based GLR hypothesis testing," *Journal of loss prevention in the process industries*, vol. 26, pp. 129–139, 2013.

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

[6] 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.

[7] 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.

[8] R. Isermann, *Fault-Diagnosis Systems*, Springer Science and Business Media, 2006.

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

[10] C. Florent, P. Borgnat, J. Tourneret, and P. Abry, "Parameter estimation for sums of correlated gamma random variables. application to anomaly detection in internet traffic.," in *ICASSP*, Las Vegas US, 2008.