A THEORETICAL APPROACH FOR INCIPIENT FAULT SEVERITY ASSESSMENT USING THE KULLBACK-LEIBLER DIVERGENCE

Jinane Harmouche[⋆†] *Claude Delpha*[⋆] *Demba Diallo*[†]

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

ABSTRACT

The limitations of statistical approaches used for Fault Detection and Diagnosis (FDD) of processes, are related to the local character of used statistics. For the sake of enhancing the detectability of incipient faults and assessing efficiently the fault level, that is the quality of operation, one should investigate the benefits a 'global' fault detection approach may provide. The Kullback-Leibler Divergence is proposed under the discipline of Statistical Process Control (SPC). A theoretical analysis is conducted to establish an analytical model relating the divergence to the fault severity amplitude. The model, when applied to a numerical example, provides an upper bound of the fault amplitude. While the usual statistics are not able to estimate the fault level, an upper bound is always necessary to guarantee a safety margin for the process.

Index Terms— Fault diagnosis, Fault modelling, Incipient fault detection, global statistical approach, Kullback-Leibler Divergence, fault severity assessment.

1. INTRODUCTION

There are extensive theoretical and experimental studies in the literature that investigate different types of Fault Detection and Diagnosis (FDD) approaches [1, 2]. Of all statistical approaches used for the FDD, Principal Component Analysis (PCA) has been successfully applied for monitoring of complex and highly dimensional processes in multiple domains [3–5]. PCA uses a large amount of historical data collected from the process to build an implicit model for the normal process behaviour. As a SPC technique, it does not require *a prior* physical full understanding of the process neither detailed theoretical studies to be made.

The present work addresses the incipient fault detection and estimation problem. The drawbacks of most statistical techniques originate from a common character. They extract a diagnostic information which actually represents only a part of the entire fault signature. This is what will be denoted by 'local' character. By analysing the FDD capabilities of the usual statistical techniques, it can be shown that limitations are strongly related to this character: (1) they show some deficiencies in the detection of incipient faults which cause small changes on the statistics of the process observations, (2) they are sensitive to the normal process variations which might obscure the fault-related information (3) the fault severity is hardly estimated using these techniques.

It can be argued that one should investigate the benefits a 'global' extraction of the fault signature may provide, in terms of enhancing the detectability of incipient faults and efficiently estimating the fault level. A 'global' extraction approach will be based on the whole fault signature in order to detect the fault, being a hidden information. The technique proposed here consists in the Kullback-Leibler Divergence (KLD) as a fault indicator and it derives from the statistical formulation of processes. This probabilistic measure was previously used for pattern recognition, anomaly and change detection [1, 6, 7]. Also, it has been shown in [8] that it is sensible to the fault level and allows an assessment of the fault severity thanks to its global character. Therefore, this paper is devoted to develop an analytical model of the divergence which will depend explicitly on the fault amplitude. The study addresses incipient faults, meaning variations ranging from 0 to 1% of signals amplitude.

The paper is organised as follows. Section 2 details the problem statement and focuses on the limitations of the usual statistical FDD approaches. Section 3 introduces the Kullback-Leibler Divergence as a 'global' fault detection criterion. A theoretical analysis of the divergence expression depending on the fault amplitude parameter is made. Section 4 is concerned with the validation of the obtained KLD analytical model. Finally, section 5 reviews the main points discussed in this work and concludes the study.

2. PROBLEM STATEMENT

The FDD is usually made by referring to the statistical moments as descriptive of probability distributions [1]. The mean, the variance, as well as the higher-order moments as skewness and kurtosis are often used separately to detect faults. This is the case of Statistical Process Control (SPC) tests used along with control charts. The on-line monitoring based on these techniques consists in evaluating, at each sampling-time, statistical tests that assess the deviation of the process from its desired values in terms of mean and variance. Multivariate control charts like the MEWMA (Multivariate Exponentially Weighted Moving Average) and the MCUSUM (Multivariate Cumulative Sum) control charts are able to detect deviations related to the process mean vector [9]. The MEWMA-CM control chart is used to detect changes on the process covariance matrix [10]. An insight of monitoring simultaneously the mean and the variance in an univariate framework was given in [11]. But it is not extended to cover the multivariate framework due to the complexity of multivariate probability distributions.

Real incipient faults affect the probability distributions in an unpredictable manner causing very slight changes along the shape of distributions. In this case, several statistical properties (e.g., mean, covariance, etc.) may change simultaneously. However, the aforementioned statistical tests are able to capture only a part of the change induced by faults. They do not reflect the actual fault level and also they may fail to detect the fault presence unless the change is quite significant.

The statistics used with projection methods such as PCA reflect instantaneous changes, and thus they are not able to give a value that indicates straightaway the fault presence, as well as its severity. The typical fault indicators used with PCA are the Hotelling $\mathbf{T^2}$ and the SPE [5].

- The $\mathbf{T^2}$ represents variations of amplitude projection of observations into the principal subspace. A relatively small change, due to the fault occurrence, will be masked by the large amount of variabilities naturally present into the principal subspace. It was argued that the principal subspace can detect only faults that affect independent variables [5]. In fact, this result underestimates the fault detection capabilities of the principal subspace, since the latter is meant to carry the most relevant information of process data.

- Correlated variables have rather to be detected with the SPE index. However, the SPE usually shows important fluctuations around the threshold whenever a fault happens. It represents variations of projections into the residual subspace, and it is sensitive to the signal level at each observation. It is not able to reflect the fault amplitude in a straightforward manner. The fault estimation based on the reconstruction principle, given in [12], is not accurate because it assumes that the SPE for faultless observations is null. However, the SPE has always a nonzero value due to the noise naturally present in the residual subspace.

Besides the $\mathbf{T^2}$, the SPE and their variants as distance discriminants, an angular-based approach is proposed in [13] in order to exploit and enhance the fault detection capability of the principal subspace. It is termed "Moving Principal Component Analysis", for which residuals are generated by comparing the direction of current principal components to the reference ones. Despite the successful applications of this approach where the local distances fail, it was shown therein that a fault is undetectable by the angular index unless it makes a meaningful change in the data correlation structure. So the detection of small faults was unsatisfactory.

From a 'global' approach's viewpoint, the fault-related information should be extracted from the whole information supplied by the data. A global fault indicator, naturally, does not reflect instantaneous variations, but it is able to capture global disparities and distortions. Instead of monitoring local parameters of probability distributions to detect changes, it becomes more appreciated to monitor the overall shape of distributions. Comparing probability distributions to their references, using informational measures, will be able to reveal small disparities caused by incipient faults; the hidden information will be totally extracted. The KLD is proposed for this objective. It allows an evaluation of the dissimilarity/divergence between two distributions.

3. INCIPIENT FAULT SEVERITY ESTIMATION

3.1. Definition

For discrimination between two continuous probability distribution functions (pdfs) $f(x)$ and $g(x)$ of a random variable x, the Kullback-Leibler Information is defined as [14]:

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

The divergence is then defined as the symmetric version of the Information:

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

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

$$KLD(f, g) = \frac{1}{2}[\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + (\mu_1 - \mu_2)^2 (\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}) - 2]. \tag{3}$$

The faults of high severity cause important changes that can affect the support of the pdfs. Only the Information can be used in this case, because the divergence requires that the pdfs share the same support. However, distortions caused by incipient faults are slight and rather invisible with an unaided eye. The divergence is preferred in this case since it gives more significant value for discrimination against the fault-free condition, and the distributions share the same support.

3.2. Analytical model derivation

A simple and light computational expression of the divergence is obtained assuming that the measurements vector 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.

Introduce the following notations:

Let's $X_{[N\times m]}$ such as $X=(\boldsymbol{x_1},...,\boldsymbol{x_j},...,\boldsymbol{x_m})=(x_{ij})_{i,j}$ is the original data matrix where $\boldsymbol{x_j}=[x_{1j}...x_{Nj}]'$ is a column vector of N measurements taken for the jth variable.

$\bar{X}_{[N\times m]}$, where $\bar{X}=(\bar{\boldsymbol{x}}_{\boldsymbol{1}},...,\bar{\boldsymbol{x}}_{\boldsymbol{j}},...,\bar{\boldsymbol{x}}_{\boldsymbol{m}})$ 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 $T=(\boldsymbol{t_1},...,\boldsymbol{t_l},...,\boldsymbol{t_m})$ 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.

The star mark (*) refers to the fault-free case.

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 fault occurrence, because we assume that 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 [13]. Then, we can write

$$\lambda_k=\lambda_k^*+\Delta\lambda_k \tag{4}$$

where $\Delta\lambda_k$ is the eigenvalue bias caused by the fault occurrence. By specializing (3) to the case considered, the divergence becomes

$$KLD(f,f^*)=\frac{1}{2}[\frac{\Delta\lambda_k^2}{\lambda_k^*(\lambda_k^*+\Delta\lambda_k)}]. \tag{5}$$

If we characterize the fault by its amplitude a, λ_k^* refers to the case $a=0$. Suppose λ_k is a function of a and is infinitely differentiable in the neighborhood of $a=0$, it stems from Taylor development of λ_k that:

$$\lambda_k=\lambda_k^*+\frac{\partial\lambda_k}{\partial a}a+\frac{1}{2}\frac{\partial^2\lambda_k}{\partial a^2}a^2+\frac{1}{3!}\frac{\partial^3\lambda_k}{\partial a^3}a^3+... \tag{6}$$

Suppose PCA is computed with the covariance matrix S, it can be shown from [15] that writing S in function of the fault parameter a gives the nth-order eigenvalue derivative as:

$$\frac{\partial^n\lambda_k}{\partial a^n}=p_k^{*\prime}\frac{\partial^n S}{\partial a^n}p_k^* \tag{7}$$

where p_k^* is the kth loading eigenvector associated to λ_k^*. Similarly to λ_k^*, p_k^* refers to the PCA's model for which $a=0$.

Then, based on (7), the KLD given by (5) will be expressed in terms of a.

So, the fault on the jth variable $\boldsymbol{x_j}$ is characterized with a multiplying factor of amplitude a affecting $\boldsymbol{x_j}$ within the sampling interval $[b,c]$. The simple fault case is considered, that is when $\boldsymbol{x_j}$ is faulty, $\boldsymbol{x_r}$ with $r\neq j$ is fault-free. We can write:

$$\boldsymbol{x_j}=\begin{bmatrix}x_{1j}\\ \vdots\\ x_{bj}\\ \vdots\\ x_{cj}\\ \vdots\\ x_{Nj}\end{bmatrix}=\begin{bmatrix}x_{1j}^*\\ \vdots\\ x_{bj}^*\\ \vdots\\ x_{cj}^*\\ \vdots\\ x_{Nj}^*\end{bmatrix}+a\times\begin{bmatrix}0\\ \vdots\\ x_{bj}^*\\ \vdots\\ x_{cj}^*\\ \vdots\\ 0\end{bmatrix}=\boldsymbol{x_j^*}+\boldsymbol{F_j} \tag{8}$$

where $\boldsymbol{F_j}=a\times[\ 0\ \ \dots\ \ x_{bj}^*\ \ \dots\ \ x_{cj}^*\ \ \dots\ \ 0\ ]'$. The sample mean of $\boldsymbol{x_j}$ is given by

$$\mu_j=\mu_j^*+a\times\frac{1}{N}\sum_{i=b}^{c}x_{ij}^*. \tag{9}$$

The jth column of the centered matrix $\bar{X}_{[N\times m]}$ is given by

$$\begin{array}{rcl}\bar{\boldsymbol{x}}_{\boldsymbol{j}} & = & \boldsymbol{x_j}-\mu_j\mathbb{1}\\ & = & (\boldsymbol{x_j^*}-\mu_j^*\mathbb{1})+(\boldsymbol{F_j}-a\times\frac{1}{N}\sum_{i=b}^{c}x_{ij}^*\mathbb{1})\\ & = & \bar{\boldsymbol{x}}_{\boldsymbol{j}}^{\boldsymbol{*}}+\bar{\boldsymbol{F}}_{\boldsymbol{j}}\end{array} \tag{10}$$

where $\bar{\boldsymbol{F}}_{\boldsymbol{j}}=\boldsymbol{F_j}-a\times\frac{1}{N}\sum_{i=b}^{c}x_{ij}^*\mathbb{1}$, $\mathbb{1}$ is a column vector of N ones. It follows from the assumption of normality of the data that the ith row of $\bar{X}$ may be assumed as a random sample of m variables drawn from a normal $\mathcal{N}(0,\Gamma)$ distribution. An unbiased estimate of the true covariance matrix Γ is given by

$$S=\frac{1}{N}\bar{X}'\bar{X}. \tag{11}$$

S can hence be written as:

$$S=\frac{1}{N}\begin{bmatrix}\bar{\boldsymbol{x}}'_{\boldsymbol{1}}\bar{\boldsymbol{x}}_{\boldsymbol{1}} & \dots & \bar{\boldsymbol{x}}'_{\boldsymbol{1}}\bar{\boldsymbol{x}}_{\boldsymbol{j}} & \dots & \bar{\boldsymbol{x}}'_{\boldsymbol{1}}\bar{\boldsymbol{x}}_{\boldsymbol{m}}\\ \vdots & & \vdots & & \vdots\\ \bar{\boldsymbol{x}}'_{\boldsymbol{j}}\bar{\boldsymbol{x}}_{\boldsymbol{1}} & \dots & \bar{\boldsymbol{x}}'_{\boldsymbol{j}}\bar{\boldsymbol{x}}_{\boldsymbol{j}} & \dots & \bar{\boldsymbol{x}}'_{\boldsymbol{j}}\bar{\boldsymbol{x}}_{\boldsymbol{m}}\\ \vdots & & \vdots & & \vdots\\ \bar{\boldsymbol{x}}'_{\boldsymbol{m}}\bar{\boldsymbol{x}}_{\boldsymbol{1}} & \dots & \bar{\boldsymbol{x}}'_{\boldsymbol{m}}\bar{\boldsymbol{x}}_{\boldsymbol{j}} & \dots & \bar{\boldsymbol{x}}'_{\boldsymbol{m}}\bar{\boldsymbol{x}}_{\boldsymbol{m}}\end{bmatrix} \tag{12}$$

Substituting $\bar{\boldsymbol{x}}_{\boldsymbol{j}}$ by its expression given in (10), and then differentiating each element of S with respect to the fault parameter a gives after all calculations:

$$\frac{\partial\bar{\boldsymbol{x}}'_{\boldsymbol{r}}\bar{\boldsymbol{x}}_{\boldsymbol{q}}}{\partial a}=0,\quad\forall r,q\neq j \tag{13}$$

$$KLD(f, f^*) = \frac{2}{N^2} \frac{\left(p_{jk} \sum_{r=1}^{m} p_{rk}\delta_r \times a + (3/2p_{jk}^2)\sigma \times a^2\right)^2}{\lambda_k^* \left(\lambda_k^* + \frac{2}{N}\left(p_{jk} \sum_{r=1}^{m} p_{rk}\delta_r\right) \times a + \frac{3}{N}(p_{jk}^2)\sigma \times a^2\right)} \quad (22)$$

$$\frac{\partial \bar{\boldsymbol{x}}_{\boldsymbol{r}}'(\bar{\boldsymbol{x}}_{\boldsymbol{j}}^* + \bar{\boldsymbol{F}}_{\boldsymbol{j}})}{\partial a} = \frac{\partial (\bar{\boldsymbol{x}}_{\boldsymbol{j}}^* + \bar{\boldsymbol{F}}_{\boldsymbol{j}})'\bar{\boldsymbol{x}}_{\boldsymbol{r}}}{\partial a} = \delta_r \quad \forall r \neq j \quad (14)$$

$$\frac{\partial (\bar{\boldsymbol{x}}_{\boldsymbol{j}}^* + \bar{\boldsymbol{F}}_{\boldsymbol{j}})'(\bar{\boldsymbol{x}}_{\boldsymbol{j}}^* + \bar{\boldsymbol{F}}_{\boldsymbol{j}})}{\partial a} = 2\delta_j + 2a\sigma. \quad (15)$$

where δ_r, δ_j and σ are constants independent of the fault parameter and given by

$$\delta_r = \sum_{i=b}^{c} (x_{ir} - \mu_r)x_{ij}^* - \frac{1}{N}\sum_{i=b}^{c} x_{ij}^* \times \sum_{i=1}^{N}(x_{ir} - \mu_r) \quad (16)$$

$$\delta_j = \sum_{i=b}^{c} (x_{ij}^* - \mu_j^*)x_{ij}^* - \frac{1}{N}\sum_{i=b}^{c} x_{ij}^* \times \sum_{i=1}^{N}(x_{ij}^* - \mu_j^*) \quad (17)$$

$$\sigma = [N - (c - b + 1)] \left(\frac{1}{N}\sum_{i=b}^{c} x_{ij}^*\right)^2 + \sum_{q=b}^{c}\left(x_{qj}^* - \frac{1}{N}\sum_{i=b}^{c} x_{ij}^*\right)^2 \quad (18)$$

δ_r, δ_j and σ are given in function of the original variables, which are not faulty, and the fault-free measurements of the variable $\boldsymbol{x_j}$ as well. The computation of these constants requires however the knowledge of the faulty interval. Therefore we may write for the first-order derivative of the covariance matrix:

$$\frac{\partial S}{\partial a} = \frac{1}{N}\begin{bmatrix} 0 & \dots & \delta_1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \delta_1 & \dots & 2\delta_j + 2a\sigma & \dots & \delta_m \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & \delta_m & \dots & 0 \end{bmatrix} \quad (19)$$

The second-order sensitivity of S with respect to the fault amplitude a is obtained by differentiating (19). We obtain:

$$\frac{\partial^2 S}{\partial a^2} = \frac{1}{N}\begin{bmatrix} 0 & & \dots & & 0 \\ & & \vdots & & \\ \vdots & \dots & 2\sigma & \dots & \vdots \\ & & \vdots & & \\ 0 & & \dots & & 0 \end{bmatrix} \quad (20)$$

The higher-order sensitivities of S $(n > 2)$ are all null and so is the case for the eigenvalue derivatives.

Writing the loading vector p_k^* as $p_k^* = \left[\begin{array}{ccc} p_{1k} & \cdots & p_{mk} \end{array}\right]'$ it follows that

$$\begin{cases} \dfrac{\partial \lambda_k}{\partial a} = p_k^{*\prime}\dfrac{\partial S}{\partial a}p_k^* = \dfrac{2}{N}\left(p_{jk}\displaystyle\sum_{r=1}^{m} p_{rk}\delta_r + p_{jk}^2\sigma \times a\right) \\ \dfrac{\partial^2 \lambda_k}{\partial a^2} = p_k^{*\prime}\dfrac{\partial^2 S}{\partial a^2}p_k^* = \dfrac{2}{N}p_{jk}^2\sigma \end{cases}$$

and thus

$$\lambda_k = \lambda_k^* + \frac{2}{N}\left(p_{jk}\sum_{r=1}^{m} p_{rk}\delta_r\right) \times a + \frac{3}{N}\left(p_{jk}^2\right)\sigma \times a^2. \quad (21)$$

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 (5), as (22).

4. MODEL VALIDATION

To validate the theoretical model of the divergence, we consider a system of m=7 variables inspired from [16] and defined as follows:

$$\begin{aligned} x_1(i) &\propto \mathcal{N}(0,1), & x_2(i) &= -x_1(i) \\ x_3(i) &= x_1(i) - x_2(i), & x_4(i) &= x_1(i) - 3x_2(i) \\ x_5(i) &= 0.5x_2(i) + x_3(i), & x_6(i) &= x_1(i) - 4x_2(i) \\ x_7(i) &= 0.2x_2(i) + x_3(i) & & \end{aligned}$$

This example is used for satisfying the theoretical assumption of the multivariate normality. We form a matrix X of N rows/samples, $X = (\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_7})$. Obviously, diagonalizing the covariance matrix of X leads to $l = 1$ eigenvalue, since $rank(X) = 1$. We get $\lambda_1^* = 52.457$. Its associated eigenvector p_1^* spans the principal subspace reduced here to an affine line: $p_1^* = [0.138\ -0.138\ 0.2761\ 0.5521\ 0.2070\ 0.6901\ 0.2484]'$. Then, we obtain $\boldsymbol{t_1^*} = \bar{X}p_1^*$, which summarizes the information contained into X. So the probability density of $\boldsymbol{t_1^*}$ is estimated as the reference distribution.

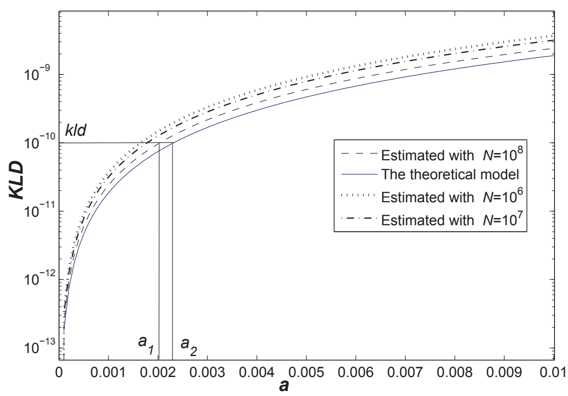

Fig. 1. Evolution of the *KLD*

Thus, the KLD evolution with respect to the fault amplitude a, will be shown through (22), as well as through the straightforward formula (2) while substituting the integral operator with a discrete summation. We set the fault interval $[b, c]$ such that the number of faulty samples consists of 1% of N. δ_r $(r = 1, ..., 7)$ and σ can hence be calculated. Then, we affect $\boldsymbol{x_4}$, chosen arbitrarily, with a fault of amplitude a (ranging from 0 to 1% as we consider incipient faults) for all samples included into $[b, c]$.

Fig.1 displays in logarithmic scale the KLD versus the fault parameter a. The continuous line refers to the theoretical model (22), and the 3 dashed ones are obtained from (2) which requires estimating the latent score's pdf for each value of a. If the sample size is large enough, the assumption of normality holds true and the analytical model gives an exact estimation of the fault amplitude. If not, however, as shown in Fig.1, the theoretical model gives an over estimation (a_2) of the actual fault amplitude (a_1) from the KLD value (kld), therefore always guaranteeing a safety margin for the faulty process.

5. CONCLUSION

While usual statistics applied for SPC were used for trending analysis, the Kullback-Leibler Divergence is able to provide a value that indicates efficiently and straightaway the fault presence as well as its severity. It has been proposed as a global statistic for detection of incipient faults (<1%). An analytical model that computes the Kullback-Leibler Divergence from the fault amplitude has been developed. The proposed theoretical approach of fault detection and diagnosis is effective and allows obtaining a safety margin according to the signal and the fault characteristics. The model has been validated with a simulated system satisfying the multivariate normality assumption. However, this assumption is not mandatory since the model still allows to give an upper bound of the fault level. In future works, this upper bound will also be theoretically verified. The FDI error probabilities will be calculated as functions of KLD.

6. REFERENCES

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

[2] M. Basseville, "On-board component fault detection and isolation using the statistical local approach," *Automatica*, vol. 34, no. 11, 1998.

[3] Y. Gu, Y. Liu, and Y. Zhang, "A selective kernel pca algorithm for anomaly detection in hyperspectral imagery," *IEEE ICASSP*, vol. 2, pp. 725–728, May 2006.

[4] C. Delpha, D. Diallo, M.E.H. Benbouzid, and C. Marchand, "Application of classification methods in fault detection and diagnosis of inverter fed induction machine drive : A trend towards reliability," *European Physical Journal of Applied Physics*, vol. 43, pp. 245–251, 2008.

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

[6] J. Silva and S. Narayanan, "Average divergence distance as a statistical discrimination measure for hidden markov models," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 3, pp. 890–906, 2006.

[7] A. Anderson and H. Haas, "Kullback-leibler divergence based anomaly detection and monotonic sequence analysis," *Vehicular Technology Conf. IEEE*, pp. 1–5, 2011.

[8] J. Harmouche, C. Delpha, and D. Diallo, "Faults diagnosis and detection using principal component analysis and kullback-leibler divergence," in *IEEE Industrial Electronics Society*, Montreal, Canada, Oct. 2012.

[9] C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon, "A multivariate exponentially weighted moving average control chart," *Technometrics*, vol. 34, no. 1, pp. 46–53, 1992.

[10] D. M. Hawkins and E. M. Maboudou-Tchao, "Multivariate exponentially weighted moving covariance matrix," *Technometrics*, vol. 50, no. 2, pp. 155–166, 2008.

[11] S. W. Cheng and K. Thaga, "Single variables control charts : an overview," *Quality and Reliability Engineering International*, vol. 22, no. 7, pp. 811–820, 2006.

[12] R. Dunia, , and S. Joe Qin, "Subspace approach to multidimensional fault identification and reconstruction," *AIChE Journal*, vol. 44, pp. 1813–1831, Aug. 1998.

[13] M. Kano, S. Hasebe, I. Hashimoto, and H. Ohno, "A new multivariate statistical process monitoring method using principal component analysis," *Comp. & Chem. Eng.*, vol. 25, pp. 1103–1113, 2001.

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

[15] N. P. Van Der Aa, H. G. Ter Morsche, and R. R. M. Mattheij, "Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem," *Electronic Journal of Linear Algebra ELA*, vol. 16, pp. 300–314, Oct. 2007.

[16] M. F. Harkat, G. Mourot, and J. Ragot, "Détection de défauts à l'aide d'une analyse en composantes principales robuste," *Sixième Conf. Int. Franc. d'Aut., CIFA*, Juin 2010.