PERFORMANCE OF JENSEN SHANNON DIVERGENCE IN INCIPIENT FAULT DETECTION AND ESTIMATION

Xiaoxia ZHANG†, Claude DELPHA†, Demba DIALLO‡

†Laboratoire des Signaux et Systèmes, CNRS - CentraleSupélec - Univ. Paris-Sud, Université Paris Saclay, Gif sur Yvette, France
‡Group of Electrical Engineering of Paris, CNRS - CentraleSupélec - Univ. Paris-Sud Sorbonne Université, Université Paris Saclay, Gif sur Yvette, France

ABSTRACT

The diagnosis (Detection, Estimation and Isolation) of incipient faults, *i.e.* faults with severity variation $< 10\%$ of the healthy signal, plays an important role in health monitoring of complex system for earliest maintenance. In this paper, Jensen-Shannon divergence (JSD) is proposed to evaluate detection and estimation of incipient fault severity in a multivariate data driven process. At first, the dimensional space is reduced, thanks to the Principal Component Analysis (PCA). Thus, the fault effect is theoretically modelled using the JSD considering Gaussian distributed signals. Then, the estimation of the fault severity is derived from this model. The fault detection performances are then computed and compared with the Hotelling's T^2 statistics. The fault estimation performances validates the theoretical modeling for incipient faults in noisy environments. An estimation error lower than 3% is obtained even for a Signal to Noise Ratio (SNR) as low as $25dB$.

Index Terms— Incipient Fault detection and estimation, Jensen-Shannon Divergence, Detection and Estimation modelling, Performances evaluation.

1. INTRODUCTION

Complex system health monitoring is becoming mandatory in different applicative sectors due to the wide development of condition based maintenance [1]. One of its main goal is to avoid abnormal stops of the process and ensure the system safety. For this purpose, fault detection and severity estimation are crucial tasks to be able to propose adapted maintenance operations. To ensure the earliest actions, the faults must be detected and diagnosed as accurately as possible at their earliest stage. Such faults type denoted incipient faults are the most difficult ones to be detected and estimated properly [2, 3]. The knowledge necessary to proceed this fault diagnosis can be based on different kind of information (physical, linguistic and data driven modelling) leading to several approaches [4, 1]. For these different approaches health monitoring is based on the evaluation of several variables named features leading to the information on the faulty behavior of the system. Detecting faults in such multivariate process is then a particular task requiring specific methodologies [5, 6].

In multivariate statistical analysis, Principal Components Analysis (PCA) is a relevant technique for analyzing and simplifying data sets. One of the main advantage of PCA is its capability to reduce the number of original variables into a principal subspace while keeping the maximum amount of information [7]. PCA-based monitoring methods are effective for handling highly dimensional, noisy and correlated data from industrial processes, and provide superior performance compared to the univariate methods. It has shown its efficiency in fault detection and estimation [8, 9, 10, 11]. For evaluating the fault detection with PCA several popular criteria can be used [8]. The Hotelling test, T^2, is the most typical and efficient criterion defined in the principal subspace [8, 12, 13]. This statistic test has shown its capability in fault detection, but its efficiency seems to be limited in noisy environments and also while the fault severity decreases. Other approaches based on the evaluation of the signals probability densities, like the Kullback-Leibler Divergence, have been proposed [14]. Nevertheless, the important variability of its value can compromise its use for instantaneous detection.

Jensen-Shannon divergence (JSD) is a sensitive technique that quantifies the Shannon entropy excess of a couple of distributions with respect to the mixture of their respective entropies [15, 16]. It has been used in several scientific areas, such as bio-informatics, genome comparison, protein surface comparison, image processing [16, 17, 18]. As this technique, based on the Shannon entropy computation, is hardly sensitive to the local fluctuations or irregularities of the probability densities, it may be suitable for incipient fault detection.

In this paper, based on the PCA framework, we propose to study the capability of JSD for incipient fault detection and fault severity estimation for a data-driven modelled process. We first propose a theoretical modeling in the case of Gaussian distributions and then evaluate the performances of the JSD in a noisy environment.

This paper is organised as follows. In section 2, the paper main contributions are summarized to highlight the benefit of this work. In section 3, the JSD technique is presented and its application in the PCA framework is described. Then a theoretical model is derived for both detection and estimation process in the incipient fault health monitoring context. In section 4, the fault detection capabilities are evaluated for an auto-regressive (AR) system. The performances are then computed and compared to those obtained with the Hotelling test. Various conditions of noise and fault severities are considered and analysed. Section 5 concludes the paper.

2. PAPER CONTRIBUTION

This paper studies the performances of JSD, for incipient fault detection and estimation in several fault severity and environmental noise conditions, using a multivariate data driven approach. These data are processed in the PCA framework to take benefit of the dimension reduction and space projection properties. In this framework, the fault effect on the JSD is theoretically derived considering Gaus-

The authors would like to thank China Scholarship Concil for funding.

sian distributed signals. Under this assumption, the model is inverted to obtain the theoretical estimation of the fault severity. The detection capability of the JSD and its performances are studied and compared to the T^2 ones in the same conditions. The superiority of JSD for incipient fault detection is then proved and its efficiency in noisy conditions is shown. The estimation error is shown to be lower than 3% even for a SNR as low as $25dB$.

3. FAULT DETECTION AND ESTIMATION

In complex systems health monitoring, it is important to detect a fault at its earliest stage. Such incipient faults are assumed to have a severity variation lower than 10% of the healthy signal [2]. This means that these slight variations can be in some cases covered by the environmental noise (negative values of the Fault to Noise Ratio as defined in [2]) and then induce missed detections. We propose here to take benefit of the Jensen-Shannon good properties for increasing the fault detection and estimation efficiency.

3.1. Jensen-Shannon Divergence for Diagnosis

3.1.1. Definition

Jensen-Shannon Divergence (JSD) is the increment of the Shannon entropy most of the time used for evaluating the distance between random graphs [15]. If we assume that f and q are two continuous probability density functions (pdfs) of a random variable x, JSD can be written as a function of the Shannon entropy (S_E) [16] and its value is denoted D_{JS} such as :

$$D_{JS}(f,q) = S_E\left[\frac{f+q}{2}\right] - \frac{S_E(f)+S_E(q)}{2} \tag{1}$$

Based on the Kullback-Leibler Information (I), JSD is defined as a symmetric operation such as :

$$D_{JS}(f||q) = \frac{1}{2}I(f||M) + \frac{1}{2}I(q||M) \tag{2}$$

where M is a mixture distribution such as $M = \frac{1}{2}(f+q)$.

and I is the Information such as : $I(f||q) = \int f(x)log\frac{f(x)}{q(x)}dx$

In this paper, we assume that the monitored system is described as a multivariate process defined with the following notations.
$X_{[N\times m]} = (\mathrm{x}_1,\cdots,\mathrm{x}_j,\cdots,\mathrm{x}_k,\cdots,\mathrm{x}_m)$ is the original data matrix, where m is the number of original variables and N is the data sample size. x_k is then the k^{th} variable as a vector $\mathrm{x}_k = [x_{1k},\cdots,x_{ik},\cdots,x_{Nk}]^{\mathrm{T}}$ with $i = [1,\cdots,N]$. x_j is the vector denoted the faulty variable within the faulty interval $[b,N]$.
$\bar{X}_{[N\times m]} = (\bar{\mathrm{x}}_1,\cdots,\bar{\mathrm{x}}_k,\cdots,\bar{\mathrm{x}}_m)$ is the centered and normalised matrix.
g denotes the fault severity amplitude while $\hat{g}$ is its estimated value. More generally, let's consider in the overall paper that the symbol (*) refers to faultless and noise-free data, ($\hat{\ }$) marks an estimated value, ($\tilde{\ }$) mention the faulty and noise-free data function and ($^{\mathrm{T}}$) is the transpose operator of a matrix.

3.1.2. Fault diagnosis in the PCA framework

For our study, PCA is used to decrease the dimensionality of the problem. Its main steps are summarized in the following. First, the sample data covariance matrix S is computed as:

$$S = \frac{1}{N-1}\bar{X}^{\mathrm{T}}\bar{X} \tag{3}$$

where each vectors of the centered and normalized matrix $\bar{X}$ are obtained as:

$$\bar{\mathrm{x}}_k = \frac{\mathrm{x}_k - u_k}{\sqrt{\sigma_k^2}} \quad (k = 1,2,\cdots,m) \tag{4}$$

where u_k and σ_k^2 are the mean and the variance of the k^{th} variable. Then the Principal Component scores matrix $T_{[N\times m]}$ is obtained as the linear transformation:

$$T_{[N\times m]} = \bar{X}_{[N\times m]}P_{[m\times m]} = (t_1,\cdots,t_k,\cdots,t_m) \tag{5}$$

where $P = (p_1,\ldots,p_l,\ldots,p_m)$ is the eigenvectors matrix of S associated to the corresponding eigenvalues $\lambda_k = [\lambda_1,\ldots,\lambda_l,\ldots,\lambda_m]$. Note that the first l component scores define the principal subspace, and the remaining $(m-l)$ define the residual subspace.

3.2. Fault detection model of JSD

For this study, let's assume that the pdf of the first l principal scores t_k denoted q_k and f_k respectively in faulty and healthy conditions are both normal distributions. Thus, we denote $q_k \sim \mathcal{N}(\mu_2,\sigma_2^2)$ and $f_k \sim \mathcal{N}(\mu_1,\sigma_1^2)$, where μ_1, μ_2 are the means and σ_1^2, σ_2^2 are the variances of f and q respectively. The mixture distribution M_k is defined as $M_k = \frac{1}{2}(f_k+q_k)$.
In the PCA's model, the variables are centered, the mean of the distribution is then supposed unchanged after the occurrence of an incipient fault [11], therefore $\mu_1 = \mu_2$. Otherwise, we know that for two normal distributions which have the same mean, the mixture distribution is unimodal [19]. In the particular case of our study for incipient faults, the modification of the pdf due to the fault is considered to be slight. So we assume that the mixture distribution M_k of the mixture signal is Gaussian distributed such as $M_k \sim \mathcal{N}(\mu_M,\sigma_M^2)$. The mean μ_M and the variance σ_M^2 of M_k are calculated respectively as (6) and (7):

$$\mu_M = \frac{1}{2}(\mu_1+\mu_2) = \mu_1 \tag{6}$$

Using the derivation of the expectation function $\mathbb{E}(.)$, we obtain:

$$\sigma_M^2 = \mathbb{E}(M_k^2) - \mathbb{E}(M_k)^2 = \frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2 \tag{7}$$

Under the normal distribution assumption, the JSD value can be derived as:

$$D_{JS}(f,q) = \frac{1}{4}\left[log\frac{\sigma_M^4}{\sigma_1^2\sigma_2^2} + \frac{\sigma_1^2+\sigma_2^2+\frac{1}{2}(\mu_1-\mu_2)^2}{\sigma_M^2} - 2\right] \tag{8}$$

In the PCA framework, σ_1^2 and σ_2^2 can be seen as a linear combination of the signal data variance given by the eigenvalue λ_k with the variance of an additive noise denoted σ_v^2 such as:

$$\sigma_1^2 = \lambda_k^* + \sigma_v^2 \qquad \sigma_2^2 = \tilde{\lambda}_k + \sigma_v^2 \tag{9}$$

where λ_k^* and $\tilde{\lambda}_k$ are respectively the eigenvalues of the k^{th} latent score in noise free healthy and faulty condition. We assume that the relation between $\tilde{\lambda}_k$ and λ_k^* is:

$$\tilde{\lambda}_k = \lambda_k^* + \Delta\lambda_k \tag{10}$$

where $\Delta\lambda_k$ is the eigenvalue bias due to the fault occurrence.
Based on equations (7), (9) and (10), the variance of M_k can be written as:

$$\sigma_M^2 = \frac{1}{2}\left(2\lambda_k^* + \Delta\lambda_k + 2\sigma_v^2\right) \tag{11}$$

Combining the equations (8), (9) and (11), the JS divergence equation is developed as:

$$D_{JS}(f||q) = \frac{1}{4}log\frac{(2\lambda_k^*+\Delta\lambda_k+2\sigma_v^2)^2}{4(\lambda_k^*+\sigma_v^2)(\lambda_k^*+\Delta\lambda_k+\sigma_v^2)} \tag{12}$$

$$D_{JS}(\Delta\lambda_k) = D_{JS}(0) + \frac{\partial D_{JS}(\Delta\lambda_k)}{\partial \Delta\lambda_k}(0)\Delta\lambda_k + \frac{1}{2}\frac{\partial^2 D_{JS}(\Delta\lambda_k)}{\partial \Delta\lambda_k^2}(0)\Delta\lambda_k^2 + \cdots \quad (17)$$

$$a_3 = \frac{\partial D_{JS}(\Delta\lambda_k)}{\partial \Delta\lambda_k} = \frac{1}{4}\left[\frac{2}{2\lambda_k^* + \Delta\lambda_k + 2\sigma_v^2} - \frac{1}{\lambda_k^* + \Delta\lambda_k + \sigma_v^2}\right] \quad (18)$$

$$a_4 = \frac{\partial^2 D_{JS}(\Delta\lambda_k)}{\partial {\Delta\lambda_k}^2} = \frac{1}{4}\left[\frac{1}{(\lambda_k^* + \Delta\lambda_k + \sigma_v^2)^2} - \frac{2}{(2\lambda_k^* + \Delta\lambda_k + 2\sigma_v^2)^2}\right] \quad (19)$$

3.3. Fault estimation model using JSD

In our study, we assume that the fault occurs on the variable x_j and affects the last $(N-b)$ observations with a fault amplitude g. The relation between $\Delta\lambda_k$ and g obtained from [11] is :

$$\Delta\lambda_k = \frac{2}{N}a_1 \times g + \frac{1}{N}a_2 \times g^2 \quad (13)$$

where a_1 and a_2 are derived in (14) and (15).

$$a_1 = p_{jk}\sum_{r=1}^{m} p_{rk}\left(\sum_{i=b}^{N}(x_{ir}^* - u_r^*)x_{ij}^*\right) \quad (14)$$

$$a_2 = 3p_{jk}^2\sum_{i=b}^{N}\left(x_{ij}^* - \frac{1}{N}\sum_{i=b}^{N}x_{ij}^*\right)^2 \quad (15)$$

Then, the theoretical equation for the fault severity estimation $\hat{g}$ is:

$$\hat{g} = \frac{-a_1 + \sqrt{a_1^2 + Na_2\Delta\lambda_k}}{a_2} \quad (16)$$

Eq.(12) can be seen as a function of $\Delta\lambda_k$ and is infinitely derivable in the neighborhood of zero. The Taylor development of D_{JS} is derived and given in (17). We get its first order derivation (18) and second order derivation (19) based on (12). Thus, limiting the Taylor equation to the first two order terms, we obtain a quadratic equation (20). Then, we can derive the approximated value of $\Delta\lambda_k$ as the solutions of equation (20). In healthy conditions, the variance change $\Delta\lambda_k$ is 0. The constants $a_3(0)$ and $a_4(0)$ can be simplified as in (21), and then we get the solution (22). The final theoretical estimation equation for $\hat{g}$ directly depending on the JSD value is obtained by combining (16) and (22):

$$D_{JS}(\Delta\lambda_k) = D_{JS}(0) + a_3(0)\Delta\lambda_k + \frac{1}{2}a_4(0)\Delta\lambda_k^2 \quad (20)$$

$$a_3(0) = 0 \qquad a_4(0) = \frac{1}{4}\left[\frac{1}{(\lambda_k^* + \sigma_v^2)^2} - \frac{2}{(2\lambda_k^* + 2\sigma_v^2)^2}\right] \quad (21)$$

$$\Delta\lambda_k = \sqrt{\frac{2D_{JS}}{a_4(0)}} \quad (22)$$

$$\hat{g} = \frac{-a_1 + \sqrt{a_1^2 + Na_2\left(\sqrt{\frac{2\hat{D_{JS}}}{a_4(0)}}\right)}}{a_2} \quad (23)$$

where $\hat{D_{JS}}$ in (23) is obtained using Monte Carlo simulations.

4. MODEL VALIDATION AND PERFORMANCE RESULTS

To validate the above derived theoretical models for JSD incipient fault detection and severity estimation, let's use a multivariate AR system such as:

$$x(i) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix} x(i-1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} u(i-1) \quad (24)$$

$$y(i) = x(i) + v(i) \quad (25)$$

where u is the correlated input

$$u(i) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix} u(i-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} w(i-1) \quad (26)$$

w is the matrix of two input uncorrelated Gaussian signals with zero mean and unit variance. u is the vector of measured inputs, and y is the output corrupted with an uncorrelated Gaussian noise with zero mean and variance σ_v^2. The vector of process variables X will be made up of the measured inputs and outputs of the process at instant value i, *i.e.* $X = \begin{bmatrix} y_1(i) & y_2(i) & u_1(i) & u_2(i) \end{bmatrix}^{\mathrm{T}}$

The faulty signal is modeled as $y_2(i) = (1+g)x_2(i) + v_2(i)$, v_2 is the additive noise with a variance σ_v^2. The fault occurs on the last 10% samples of y_2. After application of PCA we obtain 4 principal components with variances $\lambda_k^* = \{40.26; 4.9; 1.14; 0.17\}$. The first principal component (t_1) representing 86.6% of the original information is then used for the following fault detection and estimation.

4.1. Fault Detection model validation

4.1.1. Fault Detection Capability

To highlight the detection capability of the JSD, we have affected x_2 with a fault corresponding to a bias on the signal amplitude. The result is compared to the Hotelling test T^2 in the principal subspace composed of the first two principal components.

As an example, the fault detection results of T^2 and JSD are shown in Fig.1 for fault inducing a 80% bias on the signal amplitude in the last 100 samples of y_2 ($N = 1000$) and $SNR = 40dB$.

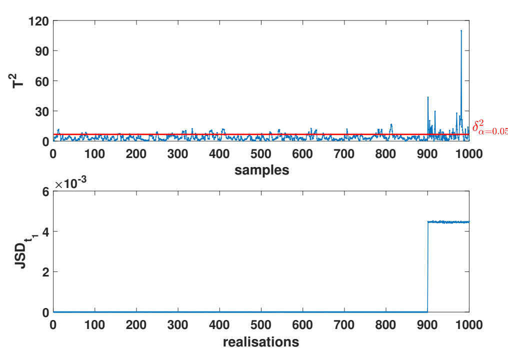

Fig. 1. Fault detection capabilities for T^2 and JSD

This example clearly shows the superiority of JSD compared to the T^2. Even if the fault severity is high, the detection with T^2 is done with numerous false alarms and missed detections. For smallest fault severities, the detection with T^2 test is obviously more difficult with poor performances while it is possible accurately with JSD.

4.1.2. Fault Detection Performances

To evaluate the detection performances for both T^2 and JSD, the receiver operating curves (ROC) are computed [20]. Several fault severities are considered and expressed in terms of Fault to Noise Ratio (FNR) [2] for $SNR = 40dB$. The performances of T^2 are depicted in Fig.2 while those of JSD are in Fig.3.

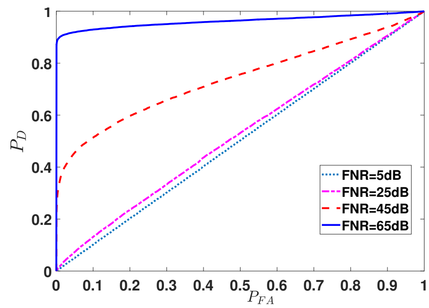

Fig. 2. Fault detection results using T^2 with SNR=40dB

Using T^2, the performances are degraded while the FNR decreases (which means lower fault severity). As defined in [2] for the considered SNR, the incipient case leads to negative or close to zero FNR values. Indeed, these results clearly confirm that the T^2 is not efficient for incipient fault detection: the detection probabilities are acceptable only in the case of large FNR values (FNR=65dB) corresponding to high fault severities.
On the opposite, JSD shows (Fig.3) excellent detection performances for small and incipient faults (FNR≥ -5dB): very good probability of detection with a very low number of false alarms. However, for lowest FNR the performances are seriously decreased.

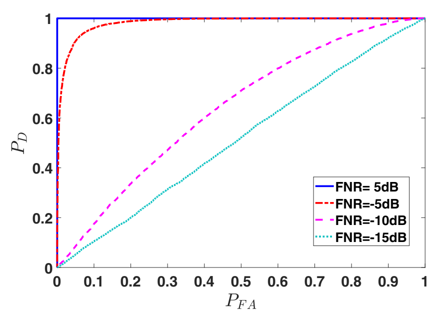

Fig. 3. Fault detection results using JSD with SNR=40dB

To highlight the noise effect on the JSD detection performances, the ROC curves are evaluated for a given fault severity $g = 0.02$ with several noise levels in the range SNR = [25 to 45] dB (see Fig.4).
JSD is efficient and has 100% detection probability for low noise levels (SNR>35dB). When the noise level increases, up to SNR=30dB, the detection performances decrease but are still acceptable. With higher noise (SNR$\leq$25dB), the performances are seriously reduced.

4.2. Fault estimation model performances

To evaluate the efficiency of the fault estimation model derived in (23), we have plotted in Fig.5 the estimated fault severity ($\hat{g}$) versus the real fault severity value (g).

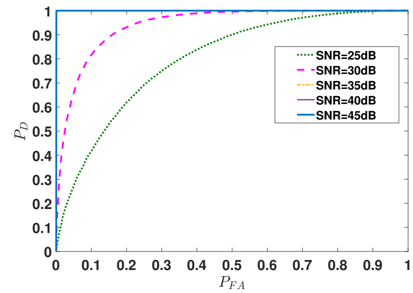

Fig. 4. Fault detection results of JSD for different SNR

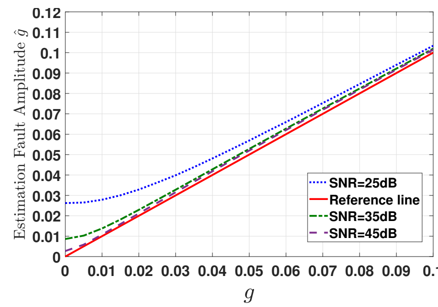

Fig. 5. Fault estimation results for different noise levels

The results highlight that the fault estimation leads to a slight overestimation of the fault severity. This overestimation provides a safety margin in a sensible fault diagnosis context. To quantify this overestimation, the accuracy of the estimation is evaluated by computing the relative error ϵ_g calculated as $\frac{\hat{g}-g}{1+g}$ (see Fig.6).

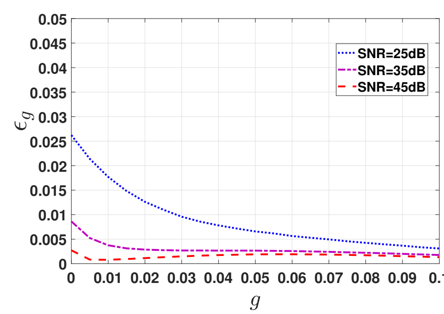

Fig. 6. Estimation relative error for different noise levels

The relative error decreases either while the fault severity or the SNR increases. For $SNR = 25dB$, the maximum error is 2.75%. For $g = 0.05$ and $SNR = 25dB$, *i.e* $FNR = -21dB$, $\epsilon_g = 2.14\%$).

5. CONCLUSION

An approach based on JSD in the PCA framework for multivariate process is proposed for incipient fault detection and severity estimation. Theoretical models are derived under the assumption of Gaussian distributed data. The fault detection performances of the JSD are compared to the T^2. The results clearly show that JSD is much more sensitive and efficient. Even in noisy conditions the performances are still acceptable. With the derived analytical model of the JSD, the fault estimation is accurate. So, JSD is a new promising solution for incipient fault diagnosis (detection and estimation) in health monitoring of complex systems.

6. REFERENCES

[1] S. Yin, X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approach for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.

[2] C. Delpha and D. Diallo, "Incipient fault detection and diagnosis : a hidden information detection problem," in *International Symposium on Industrial Electronics (ISIE 2015)*. Rio de Janeiro, Brasil: IEEE, June 3-5 2015, pp. 837 – 842.

[3] E. Chaing, L. Russel and R. Braatz, *Fault detection and diagnosis in industrial systems*. Springer, London, 2001.

[4] R. Isermann, "Process fault detection based on modeling and estimation methods-a survey," *Automatica*, vol. 20, no. 4, pp. 387–404, July 1984.

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

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

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

[8] Y. Tharrault, G. Mourot, J. Ragot, and D. Maquin, "Fault detection and isolation with robust principal component analysis," *International Journal of Applied Mathematics and Computer Science*, vol. 18, no. 4, pp. 429–442, 2008.

[9] 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.

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

[11] ——, "Incipient fault amplitude estimation using KL divergence with a probabilistic approach," *Elsevier Journal on Signal Processing*, vol. 120C, pp. 1–7, March 2016.

[12] H. Wang, Z. Song, and P. Li, "Fault detection behavior and performance analysis of principal component analysis based process monitoring methods," *Industrial & Engineering Chemistry Research*, vol. 41, no. 10, pp. 2455–2464, September 2002.

[13] H. Hotelling, "Analysis of a complex of statistical variables into principle components," *Journal of education psychology*, vol. 24, pp. 417–441, 1933.

[14] 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, no. 1, pp. 278–287, January 2014.

[15] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Trans. Information Theory*, vol. 37, pp. 145–151, 1991.

[16] P. Sánchez-Moreno, A. Zarzo, and J. Sánchez-Dehesa, "Jensen divergence based on fisher's information," *Journal of Physics A: Mathematical and Theoretical*, vol. 45, no. 12, 2012.

[17] W. Yang, H. Song, X. Huang, X. Xu, and M. Liao, "Change detection in high-resolution sar images based on jensen shannon divergence and hierarchical markov model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 3318–3327, September 2014.

[18] C. Atae-Allah, J. Gomez-Lopera, P. Luque-Escamilla, R. Martinez-Aroza, and R. Roman-Roldan, "Image segmentation by jensen-shannon divergence. application to measurement of interfacial tension," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 3, Barcelona, Spain, September 3 - 7 2000.

[19] I. Eisenberger, "Genesis of bimodal distributions," *Technometrics*, vol. 6, no. 4, pp. 357–363, 2012.

[20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861 – 874, 2006.