Incipient fault detection and estimation based on Jensen–Shannon divergence in a data-driven approach

Xiaoxia Zhang, Claude Delpha, Demba Diallo

Incipient Fault Detection and Estimation Based on Jensen-Shannon Divergence in a Data-Driven Approach

Xiaoxia ZHANG*, Claude DELPHA*, Demba DIALLO†

*Laboratoire des Signaux et Systemes (L2S),
CNRS - CentraleSupelec - Univ. Paris Sud - Université Paris-Saclay,
3, Rue Joliot Curie, Gif Sur Yvette, France

†Group of Electrical Engineering of Paris (GeePs),
CNRS - CentraleSupelec - Univ. Paris Sud - Sorbonne Univ. - Université Paris Saclay,
11, Rue Joliot Curie, Gif Sur Yvette, France

Abstract

Most data-driven diagnosis methods that are designed to detect faults, rely on measuring the mean and variation shifts. However, for incipient fault detection, these statistical criteria are slightly varying and are difficult to be accurately evaluated to reach good performances. Indeed, such faults are more likely to induce slight changes on the probability distribution rather than particular parametric changes. Therefore, the Jensen-Shannon Divergence(JSD), characterized by high sensitivity in measuring minor changes between probability distributions, is proposed in this paper. Its efficiency for detection and estimation is theoretically studied and validated by simulated data considering an auto-regressive (AR) system designing a multivariate data-driven process. The superior detection performances are demonstrated and compared with other more traditional statistical tests such as the Hotelling's T^2 and the Squared Prediction Error (SPE) in the Principal Component Analysis (PCA) framework. Minor crack detection based on eddy-currents testing (ECT) experimental data are evaluated to highlight the performances of our proposal. The results show that JSD can detect minor cracks ($0.01mm^2$ to $0.04mm^2$) which were not possible when using the baseline impedance signal measurement. For the fault severity estimation, the accuracy of the theoretical model derived for Gaussian distributed signals is shown with an AR system. The maximum relative estimation error obtained in the worst faults severity conditions is smaller than 2.75% when the Signal to Noise Ratio (SNR) is larger than 25dB and smaller than 2.15%

when the Fault to Noise Ratio (FNR) is larger than −21dB. Application for the fault severity estimation on the ECT data validates the effectiveness of this fault estimation model.

Keywords: Incipient fault, Fault detection and estimation, Jensen Shannon Divergence, Data-driven process, Principal Component Analysis.

1. Introduction

Incipient fault diagnosis plays a key role in the automation of inspection procedure and minimization of maintenance activities and costs [1, 2]. An incipient fault is mainly characterized by its slowly developing behavior and its barely noticeable effects [3, 4, 5]. Besides this type of defect is difficult to detect, it is dangerous for a system if its severity is underestimated. The early detection of the incipient fault can help to schedule preventive maintenance and prevent more serious failures [6]. For quantifying the fault severity, the estimation of the incipient fault is necessary.

The existing methods of fault detection and estimation can be generally classified as model-based and data-driven-based solutions [7, 8]. The model-based approach is related to the theoretical derivation of the system behavior. Its performances widely depend on the accuracy of the mathematical model which is degraded due to uncertainties, evolution of operating conditions, environmental changes, etc [9]. In contrast to the model-based approach, the process-history-based methods are based on the availability of a sufficient amount of historical process data to perfectly describe the process behavior using well-chosen descriptive features [5, 10, 11]. Basically, techniques that are often used for data-driven approaches are linked to the evaluation of statistical moments with order 1 to 4 (sample mean, variance, skewness, and kurtosis), Hotelling's T^2, squared prediction error (SPE), Cumulative Sum (CUSUM), ... [7, 12, 13]. These methods are efficient in measuring the parameter change of the data but are not accurate enough for incipient fault detection. These later seem more likely to cause a change of the probability distribution rather than obvious parameter change [14]. Particular techniques based on specific distance measures have then to be considered to detect such faults [15, 16]. Jensen-Shannon divergence (JSD) is a particular symmetrical operation of Kullback-Leibler information. It is a well-known tool for detecting the dissimilarities between probability distributions [17, 18]. It has been used in many domains such as image processing, text categoriza-

tion, and subject recognition. Its superior change detection capability has been well observed in [19, 20, 21, 22] for particular contexts. Nevertheless, to the best of our knowledge, the efficiency of this technique has never been considered in the particular working conditions of incipient fault that can be masked by the noise environment.

In this paper, JSD is proposed to detect incipient fault for different noise levels in a multivariate data-driven process. Firstly, a theoretical model is developed to prove the detection efficiency of the technique. As a preprocessing and feature extraction method, Principal Component Analysis (PCA), which effectiveness in fault detection and estimation is well known [23, 24, 25, 26], is used to reduce the dimensionality of the processed variables [27] and extract the most informative feature. Meanwhile, the maximum information is kept in this new feature space and the probability density functions of the system in healthy and faulty conditions can be derived. The detection performances are evaluated considering JSD but also typical statistics algorithms (Hotelling's T^2 and SPE) with well-proved detection capabilities in process health monitoring [7, 26, 28, 29]. These performances are shown using a simulated AR process and compared to each other. To experimentally validate this detection step, minor material cracks based on Eddy Current Testing (ECT) data are considered.

The existence of material crack is harmful to the safety of an industrial system particularly in the transportation industry (planes, rockets, trains, vehicles, $\cdots$). In order to detect these cracks, a regular inspection of the system is required during all the production and maintenance process. Non-destructive evaluation (NDE) technique can be used to test the inside defects of materials and structures without affecting its operating conditions [30, 31]. For conductive metallic materials, Eddy current based techniques such as Pulsed Eddy current (PEC) or Eddy Current Testing (ECT) are the main typical ones that are most often used as non-destructive testing methods for crack detection [32, 33, 34]. They are based on the principle that the inhomogeneity due to the defect will cause measurement shifts compared to the initial healthy conditions. To be applied in high-dimensional data conditions, fault feature extraction methods such as Principal Component Analysis (PCA) [35], Independent Component Analysis (ICA) [36], or Fisher Discriminant Analysis (FDA) [37] have been applied for the cracks diagnosis. Most of these works are focused on large cracks detection. This implies a significant visible signature due to the presence of the fault in the eddy current signals with a very low influence of the environmental noise. However,

in the case of minor cracks, which can be considered as incipient faults, the measurement variation due to the fault is weak. The fault signature compared to the roughness leads to a measured signal that does not clearly indicate the presence of the crack. Particularly, the measurement accuracy is largely affected by external measurement environment, the roughness of the surface, internal and sensor noise sensitivity, $\cdots$ [33, 38]. Thus, the detection and diagnosis for these incipient cracks in a high noise level are more tricky. In this paper, Jensen-Shannon divergence is proposed and used for incipient cracks detection in a high noise level based on the experimental ECT data. With our proposal, we show that it is possible to detect smaller cracks than those that have already been evaluated in the literature to the best of our knowledge.

The second aim of this paper concerns the fault severity estimation. Based on the JSD, we propose a theoretical model derived for Gaussian distributed data. The efficiency of this estimation process is validated using the simulated AR process and the experimental ECT data. As the incipient faults lead to very low fault severities they can be masked by the environmental noise of the process. Then, the accuracy of our estimation results is discussed for several operating conditions corresponding to different Signal to Noise Ratio (SNR) and Fault to Noise Ratio (FNR). With our proposal this estimation is, to the best of our knowledge, the first that have been obtained for those crack size in noisy environment. Offering a slight overestimation of the fault severity and a very low error rate, this proposal allows to have a valuable safety margin in the health monitoring process.

This paper is organized as follows. The detailed incipient fault diagnosis process and main notations are described in Section 2. Section 3 and 4 respectively presents the derived incipient fault detection and estimation models with the evaluation of their performances and their application to ECT data. The last section 5 concludes the paper.

2. Incipient Fault Diagnosis Process

2.1. Main notations and fault model

Let's first define the main notations used throughout the paper:

- $\mathbf{X}_{[N\times m]} = (\mathbf{x}_1, \cdots, \mathbf{x}_j, \cdots, \mathbf{x}_k, \cdots, \mathbf{x}_m)$ represents original data matrix, in which m is the number of variables and N is the sample size. $\mathbf{x}_k$ is the k^{th} vector such as $\mathbf{x}_k = [x_{1k}, \cdots, x_{ik}, \cdots, x_{Nk}]^\intercal$, where $i = [1, \cdots, N]$

is the sample number. x_{ik} is then the i^{th} data sample of the k^{th} variable. $\mathbf{x}_j$ denotes the variable affected by the fault occurred within the sampling interval $[b, N]$, $(^\intercal)$ is the transpose matrix operator.

- $\mathbf{S}$ is the sample data covariance matrix,
- $\bar{\mathbf{X}}_{[N\times m]}$ is the centered and normalized matrix.
- l corresponds to the dimension of the principal subspace, leading to the l first principal components. The $m-l$ remaining ones give the dimension of the residual subspace.
- $\mathbf{P}_{[m\times m]}$ is the eigenvectors matrix (also denoted the loadings matrix) of $\mathbf{S}$ associated to the eigenvalues $\mathbf{\Lambda}$, such as $diag(\mathbf{\Lambda}) = [\lambda_1, \cdots, \lambda_k, \cdots, \lambda_m]$.
- $D_{JS}(.)$ denotes Jensen-Shannon Divergence, $D_{JS_{An}}$ is the approximated theoretical equation of JSD derived for Gaussian distributed data.
- l is the crack length, and d is crack depth.
- pdf denotes the probability density function.
- P_{FA} refers to the probability of false alarm, and P_D is the probability of detection.
- The mark $(*)$ refers to faultless and noise-free data, mark $(\tilde{})$ mentions the faulty and noise-free data function, $(\breve{})$ marks the faulty and noisy data, $(\hat{})$ marks the estimated function.

For our work, the fault model is based on the assumption that the fault severity is constant. The fault occurres at instant time b on the j^{th} variable, whereas its last $(N-b)$ observations are affected by the fault.

A white Gaussian signal $\mathbf{v}_k \sim \mathcal{N}(0, \sigma_v^2)$ is used to simulate the noise in the environment. The noise matrix is then $\mathbf{V}_{[N\times m]} = (\mathbf{v}_1, \cdots, \mathbf{v}_k, \cdots, \mathbf{v}_m)$, where $\mathbf{v}_j = [v_{1j}, \cdots, v_{ij}, \cdots, v_{Nj}]^\intercal$ is the noise vector of N samples that affecting the j^{th} variable. Thus, the faulty variable $\mathbf{x}_j$ can be written as:

$$\mathbf{x}_j = \mathbf{x}_j^* + \mathrm{F}_j + \mathbf{v}_j \tag{1}$$

where F_j is the fault and its structure is denoted as:

$$\mathrm{F}_j = g \times [0 \ldots 0 \ldots x_{bj}^* \ldots x_{Nj}^*]^\intercal \tag{2}$$

where g is the fault severity amplitude.

2.2. Diagnosis process

The diagnosis procedure proposed for fault detection and estimation of a multivariate data-driven process in the Principal Component Analysis framework is shown in Fig.1.

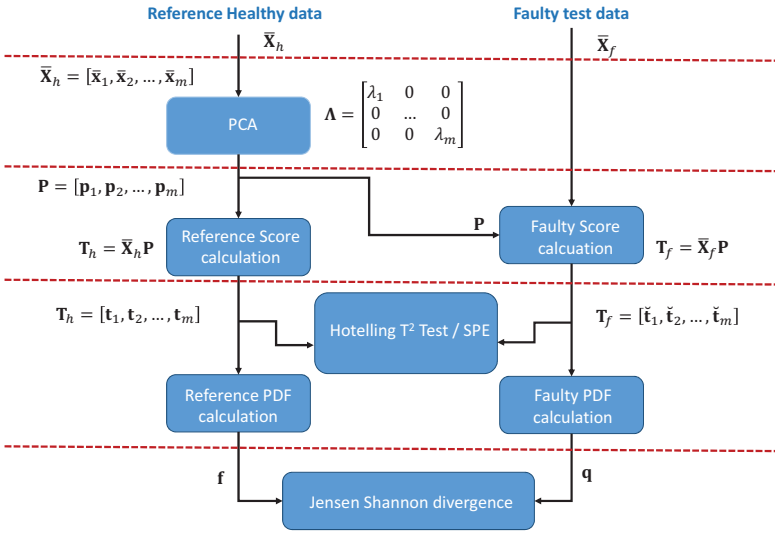


Figure 1: Proposed fault diagnosis procedure

The main operations of the diagnosis procedure based on JSD, T^2 and SPE are described in the following items:

- The input data of PCA are first preprocessed and mean-centered. The PCA model is then established based on the reference healthy data. So, the eigenvalues and eigenvectors matrices $\mathbf{\Lambda}$ and $\mathbf{P}$ are derived respectively.
- The reference score $\mathbf{T}_h$ and the faulty score $\mathbf{T}_f$ are obtained from the data matrices using the same eigenvector matrix $\mathbf{P}$.
- The probability densities of each latent score of $\mathbf{T}_h$ and each latent score of $\mathbf{T}_f$ are estimated using a normal kernel estimator.
- The diagnosis using Jensen-Shannon divergence is based on the pdfs of the first principal component scores of $\mathbf{T}_h$ and $\mathbf{T}_f$. The diagnosis using Hotelling's T^2 is based on the principal component scores in the principal subspace, and SPE uses the scores of the residual one.

2.3. Materials and methods for our proposal

2.3.1. Principal Component Analysis

In fault detection process for multivariate systems, Principal Component Analysis (PCA) has been widely used as preprocessing (filtering and dimension reduction) or feature extraction (fault feature, component selection, ...) [24, 25, 26, 29, 30, 39]. In our study, PCA is used for dimension reduction and component selection. Its major steps can be summarized as follows.

First, the covariance matrix $\mathbf{S}$ of sample data can be defined as:

$$\mathbf{S} = \frac{1}{N-1}\bar{\mathbf{X}}^{\intercal}\bar{\mathbf{X}} \tag{3}$$

where $\bar{\mathbf{X}}_{[N\times m]}$ is the centered and normalized matrix. Each vector of this matrix $\bar{\mathbf{X}}$ is written as:

$$\bar{\mathbf{x}}_k = \frac{\mathbf{x}_k - \mu_k}{\sqrt{\sigma_k^2}} \tag{4}$$

where μ_k and σ_k^2 are the mean and the variance for the k^{th} variable ($k = 1, 2, \cdots, m$). The Principal Component scores matrix $\mathbf{T}_{[N\times m]}$ can be determined using the linear transformation:

$$\mathbf{T}_{[N\times m]} = \bar{\mathbf{X}}_{[N\times m]}\mathbf{P}_{[m\times m]} = (\mathbf{t}_1, \cdots, \mathbf{t}_k \cdots, \mathbf{t}_m) \tag{5}$$

where $\mathbf{P} = (\mathbf{p}_1, \cdots, \mathbf{p}_l, \cdots, \mathbf{p}_m)$ is the eigenvectors matrix of $\mathbf{S}$ associated to the corresponding eigenvalues $\lambda_1, \cdots, \lambda_l, \cdots, \lambda_m$.

The principal subspace is defined by the first l principal components, and the residual one is determined by the remaining $(m-l)$ components. In our paper, the number of principal components (PCs) l is obtained calculating the Cumulative Percent of Variance (CPV) [39, 40].

$$\text{CPV}(l) = \frac{\sum_1^l \lambda_k}{\sum_1^m \lambda_k} \geq 90\% \tag{6}$$

The number of PCs cumulatively contributing to more than 90 percent data variance will lead us to the principal subspace.

2.3.2. Hotelling's T^2 and SPE

Typical detection indices are distance-based, aiming at evaluating how much a new observation is away from each of the subspaces. The Hotelling's T^2 is the typical statistic test based on the principal subspace, and SPE is the

typical statistic criterion based on the residual subspace. The observations of T^2 instant i is:

$$T^2 = \sum_{k=1}^{l} \frac{t_{ik}^2}{\lambda_k} \tag{7}$$

The threshold for T^2 can be approximated by:

$$T_{l,\alpha}^2 = \frac{l(N^2-1)}{N(N-l)} \mathcal{F}_{l,N-l,\alpha} \tag{8}$$

where $\mathcal{F}_{l,N-l,\alpha}$ is the Fisher distribution with two degrees of freedom, l and $N-l$.

The observations of SPE at instant i is:

$$SPE = \sum_{k=l+1}^{m} t_{ik}^2 \tag{9}$$

The theoretical threshold of SPE at significance level α is:

$$\delta_\alpha^2 = \xi \chi_{h,\alpha}^2 \tag{10}$$

where $\xi = \gamma_2/\gamma_1$, $h = integer(\gamma_1^2/\gamma_2)$, $integer(o)$ is the integer value of o and $\chi_{h,\alpha}$ is the Chi-square distribution with h degrees of freedom. The constant γ_c is calculated as $\gamma_c = \sum_{k=l+1}^{m} \lambda_k^c$ and λ_k^c is the k^{th} eigenvalue to the c^{th} power (c={1,2}).

2.3.3. Jensen-Shannon divergence

Jensen Shannon divergence (JSD) is a sensitive technique based on Shannon entropy excess of a couple of distributions in regard to the mixture of their respective entropies and without assumptions on their types [41]. Considering $\mathbf{f}(t)$ and $\mathbf{q}(t)$ are continuous probability density functions (pdfs) corresponding to the random variable t, JSD is the increment of the Shannon entropy. It is defined by a function of the Shannon entropy (S_E) and its value is denoted D_{JS} such as:

$$D_{JS}(\mathbf{f}, \mathbf{q}) = S_E \left[\frac{\mathbf{f}+\mathbf{q}}{2} \right] - \frac{S_E(\mathbf{f}) + S_E(\mathbf{q})}{2} \tag{11}$$

The JSD can be also written as the symmetric operation of the Kullback-Leibler information theoretical function denoted as:

$$D_{JS}(\mathbf{f}, \mathbf{q}) = \frac{1}{2} I(\mathbf{f}||\mathrm{M}) + \frac{1}{2} I(\mathbf{q}||\mathrm{M}) \tag{12}$$

where $\mathrm{M} = \frac{1}{2}(\mathbf{f} + \mathbf{q})$ is a mixture distribution, I is the Kullback-Leibler information [42] defined as:

$$I(\mathbf{f}||\mathbf{q}) = \int \mathbf{f}(t) log \frac{\mathbf{f}(t)}{\mathbf{q}(t)} dt \tag{13}$$

As the divergence has no closed-form, the integral function (13) is numerically calculated using the Monte-Carlo approximation.

Otherwise, when the divergence value is theoretically equal to zero, it means that the two considered probability density functions are exactly the same: the reference one and the tested one are both obtained in healthy condition. In real life applications and experimentally measured data, these two distributions won't be exactly the same. Due to the environmental nuisances, the two functions will be slightly different: this will lead to a low non-zero divergence due to the random noises. To proceed to an efficient decision and accurately highlight whether the incipient crack exists, the detection must be done with respect to a given threshold. This threshold is settled by evaluating the JSD on the pdfs for the process data in healthy conditions. Below this threshold, a minimized number of false alarms must be considered meaning that there will be a reduced number of false information about the presence of an incipient crack. At the same time, a maximized number of true detection of a considered incipient crack should be obtained. The evolution of this threshold is then carried out with the receiver operating curve representation highlighting the detection performances.

3. Incipient Fault Detection

3.1. Fault detection theoretical model

As mentioned in the previous section 2, PCA is first applied as the preprocessing method and feature extraction method. Let's denote $\mathbf{q}_k$ and $\mathbf{f}_k$ the pdfs of the first l principal score obtained in faulty and healthy conditions respectively. For Gaussian distributed signals, we can assume that the principal scores are Gaussian distributed. Then, $\mathbf{f}_k \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathbf{q}_k \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where μ_1, μ_2 are the means and σ_1^2, σ_2^2 are the variances. The mixture distribution M_k is computed by $\mathrm{M}_k = \frac{1}{2}(\mathbf{q}_k + \mathbf{f}_k)$.

In the PCA's model, the mean of the distribution is supposed unchanged after the incipient fault occurrence [25, 43] and described in (14).

$$\mu_1 = \mu_2 \tag{14}$$

And for two normally distributed functions sharing the same mean, the combined distribution is unimodal [44]. In the particular case of incipient faut detection, the change on the pdf induced by the incipient fault is very slight. Thus, we can consider the assumption that the mixture distribution M_k is still normally distributed, such as $\mathrm{M}_k \sim \mathcal{N}(\mu_{\mathrm{M}}, \sigma^2_{\mathrm{M}})$. The mean μ_{M} and the variance σ^2_{M} of M_k can be calculated in (15) and (16):

$$\mu_{\mathrm{M}} = \frac{1}{2}(\mu_1 + \mu_2) \tag{15}$$

$$\sigma^2_{\mathrm{M}} = \mathbb{E}(\mathbf{t}^2) - \mathbb{E}(\mathbf{t})^2 = \frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2 \tag{16}$$

The expression of the JSD under the Gaussian distribution assumption is:

$$D_{JS_{An}}(\mathbf{f}, \mathbf{q}) = \frac{1}{4}[log\frac{\sigma^4_{\mathrm{M}}}{\sigma_1^2\sigma_2^2} + \frac{\sigma_1^2 + \sigma_2^2 + \frac{1}{2}(\mu_1 - \mu_2)^2}{\sigma^2_{\mathrm{M}}} - 2] \tag{17}$$

In the latter, the variances σ_1^2 and σ_2^2 in (17) can be rewritten considering the component data variance and the additive noise ones such as:

$$\sigma_1^2 = \lambda_k^* + \sigma_v^2 \qquad \sigma_2^2 = \tilde{\lambda}_k + \sigma_v^2 \tag{18}$$

where λ_k^* and $\tilde{\lambda}_k$ are the eigenvalues of the k^{th} latent score under healthy and faulty conditions without noise, respectively. σ_v^2 is the variance of the additive noise.

We assume that there is a relation between $\tilde{\lambda}_k$ and λ_k^* :

$$\tilde{\lambda}_k = \lambda_k^* + \Delta\lambda_k \tag{19}$$

where $\Delta\lambda_k$ is the eigenvalue bias due to incipient the fault occurrence.

Combining (16), (18) with (19), the variance of M_k is obtained:

$$\sigma^2_{\mathrm{M}} = \frac{1}{2}(2\lambda_k^* + \Delta\lambda_k + 2\sigma_v^2) \tag{20}$$

Based on (14), (16) and (20), equation (17) can be transformed into:

$$D_{JS_{An}}(\mathbf{f}, \mathbf{q}) = \frac{1}{4}log\frac{(2\lambda_k^* + \Delta\lambda_k + 2\sigma_v^2)^2}{4(\lambda_k^* + \sigma_v^2)(\lambda_k^* + \Delta\lambda_k + \sigma_v^2)} \tag{21}$$

The obtained equation (21), clearly shows the relation between $\Delta\lambda_k$ and the JSD value.

3.2. Incipient fault detection performances results and discussion

The traditional multivariate AR system with Gaussian distributed data is used as an example to evaluate the detection performance of T^2, SPE and JSD. We consider:

$$\mathbf{x}(i) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix}\mathbf{x}(i-1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}\mathbf{u}(i-1) \tag{22}$$

$$\mathbf{y}(i) = \mathbf{x}(i) + \mathbf{v}(i) \tag{23}$$

where $\mathbf{u}$ is the correlated input

$$\mathbf{u}(i) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix}\mathbf{u}(i-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix}\mathbf{w}(i-1) \tag{24}$$

$\mathbf{w}(i) = [w_1(i) \quad w_2(i)]^\intercal$ is the input vector built with two uncorrelated Gaussian signals with zero mean and unit variance. $\mathbf{u}(i) = [u_1(i) \quad u_2(i)]^\intercal$ is the measured input vector. $\mathbf{y}(i) = [y_1(i) \quad y_2(i)]^\intercal$ is the output vector and $\mathbf{v}(i) = [v_1(i) \quad v_2(i)]^\intercal$ is the noise vector which built with two uncorrelated Gaussian noise signals with zero mean and variance σ_v^2. The matrix $\mathbf{X}$ is formed with the AR measured inputs and outputs at instant i with $i = 1 \dots N$, i.e. $\mathbf{X} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \mathbf{u}_1 \quad \mathbf{u}_2]^\intercal$. Where each vectors $\mathbf{y}_c$ and $\mathbf{u}_c$ can be respectively written as $\mathbf{y}_c = [y_c(1), \dots, y_c(i), \dots, y_c(N)]^\intercal$ and $\mathbf{u}_c = [u_c(1), \dots, u_c(i), \dots, u_c(N)]^\intercal$ with $c = \{1, 2\}$.

After the application of PCA, we obtain four principal components with the following eigenvalues $diag(\mathbf{\Lambda}) = [40.26, 4.9, 1.14, 0.17]$. The corresponding percent of variance for the four PCs are then $[86.64, 10.54, 2.45, 0.37]$. So, the first PC contains 86.64% of the original data overall information and the CPV of the first two is 97.18%. With respect to equation (6), the principal subspace is then composed of the first two principal components and the residual one with the last two components. So, for this validation study, the detection process using T^2 is then based on the first two principal component scores and SPE is evaluated using the last two components. For our proposal using JSD, we will focus only on the first principal component which contains a sufficiently significant part of the data information (86.64%).

Based on the mentioned fault model, assuming that the fault affects the last 10% samples of $\mathbf{y}_2$, we have $y_2(i) = (1+g)x_2(i)+v_2(i)$. Thus, to highlight the detection capabilities, we have affected $\mathbf{x}_2$ with a fault corresponding to 80% bias on the signal amplitude in the last 100 samples of $\mathbf{x}_2$ with $N = 1000$

and $SNR = 40dB$. For the T^2 test and SPE, the detection is evaluated on each sample i but for JSD, the results are obtained by doing 900 realisations of the healthy conditions and 100 realisations of the faulty ones. The results are displayed in Fig.2. The obtained detection results can then be compared.

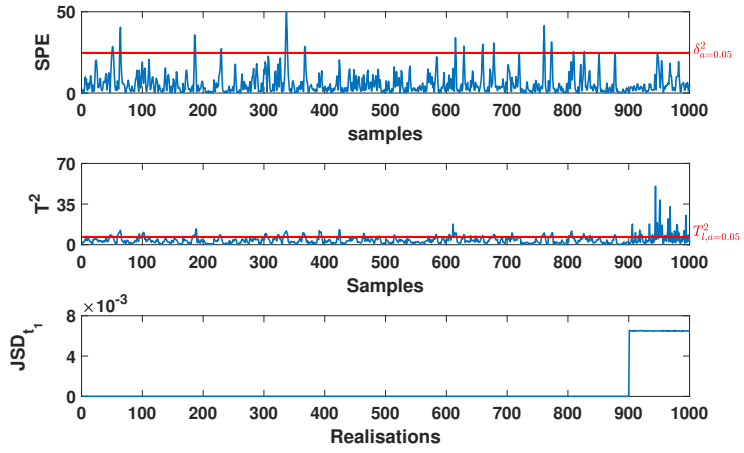


Figure 2: Fault detection performances of JSD, T^2 and SPE

It is noted from this example that JSD is superior to the other criteria T^2 and SPE. Even when the fault severity is very high, the detection with T^2 and SPE reveals numerous false alarms and missed detections. For a given $P_{FA} = 0.05$ the P_D evaluated for T^2, SPE and JSD are respectively 0.63, 0.07 and 1. Indeed for smallest fault severities, T^2 and SPE will exhibit poorer performances compared to JSD.

According to equation (21), the detection performances of the proposed methodology are largely affected by the evolution of the noise level. To highlight this purpose, we propose to evaluate these performances considering the noise level compared to the original signal power (signal to noise ratio i.e. SNR) or the fault power (fault to noise ratio i.e. FNR) defined as:

$$FNR = 10 \cdot \log_{10} \frac{\sigma_f^2}{\sigma_v^2} \quad (25)$$

$$SNR = 10 \cdot \log_{10} \frac{\sigma_s^2}{\sigma_v^2} \quad (26)$$

where σ_f^2 is the fault power, σ_s^2 is the signal power and σ_v^2 is the noise power.

Considering that the noise can mask the incipient fault, the comparative study is first done by setting the SNR to 40dB and varying the fault severity. The detection performances results for the incipient fault of T^2, SPE, and JSD for different FNR are displayed in Fig.3.

It is clear that the detection performances of T^2 and SPE are widely degraded with the decrease of FNR (i.e. lower fault severity can be obtained at a given noise level).

It is demonstrated that T^2 and SPE are totally inefficient for incipient fault detection. Their detection performances are only acceptable for large FNR values (e.g. FNR is 65dB) that is corresponding to large fault severities.

Conversely, JSD shows excellent performances for incipient faults detection. The probability of detection is $P_D = 1$ even when $FNR = 5$dB. For lower fault severities the performances are still good for FNR$=-5$dB but for FNR values lower than -10dB the detection is more tedious.

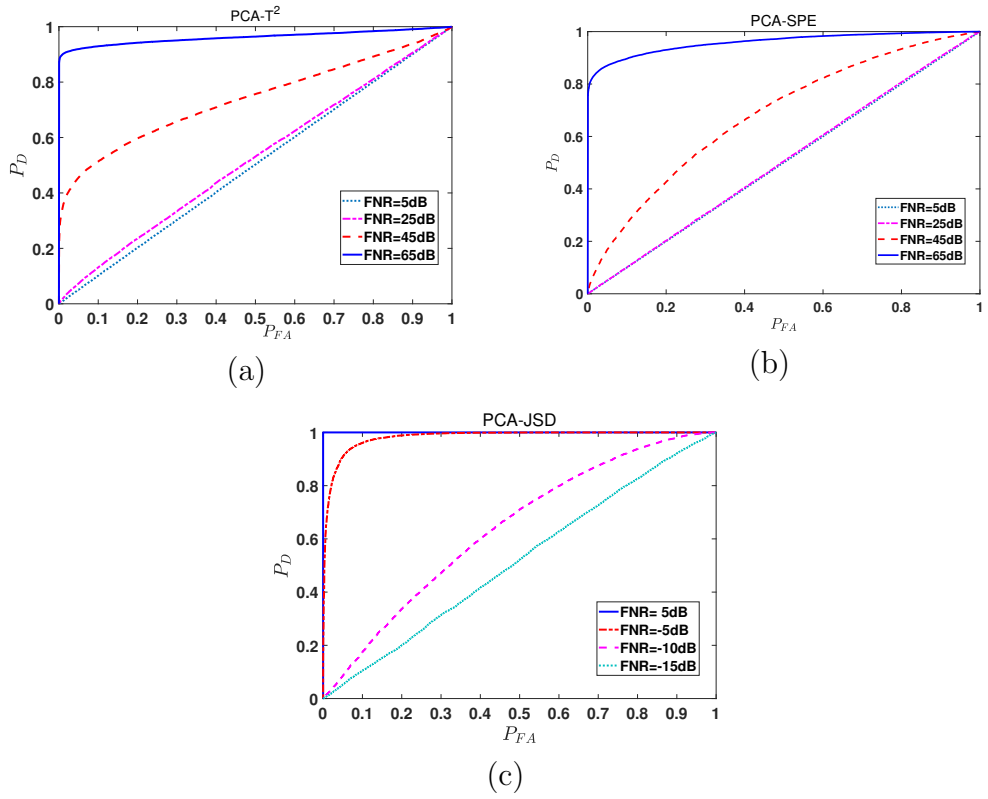


Figure 3: Fault detection performances considering different FNR with SNR=40dB for: (a) T^2, (b) SPE and (c) JSD

In order to show the impact of noise factors on the detection capabilities of JSD, we consider a fault severity $g = 0.02$. The detection results versus different SNR are given in Fig.4. JSD has high efficiency with 100% detection capability for the low noise levels (SNR>35dB) with a very low false alarm probability. For the high noise levels (SNR<30dB), the detection performances of JSD are affected by the noise. The fault detection performances of JSD decrease distinctly along with the SNR decreases (the noise levels increase).

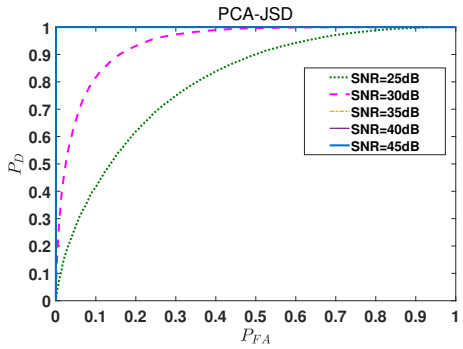


Figure 4: Fault detection results of JSD for different SNR, $g = 0.02$

JSD is then efficient for the detection of incipient faults but its performances are clearly affected when the noise level is very important compared to the fault severity one.

3.3. Application to nondestructive incipient cracks detection

Material crack detection becomes more and more important for industrial system and structures. The Eddy Current Testing (ECT) technique is a widely used measurement approach. It is based on the measurement of the magnetic field effects (induced voltage and current flowing in an excitation coil) [31]. In Fig.5 we show an example of the change induced on the eddy currents trajectory in a conductive material with the presence of a surface crack.

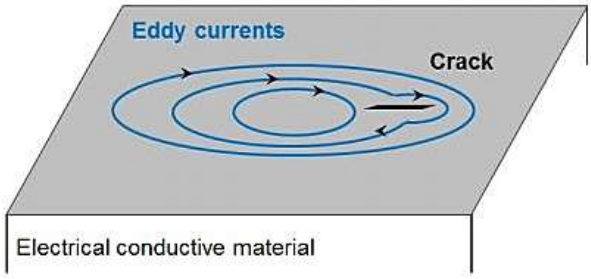


Figure 5: Effect of the presence of a crack on Eddy currents trajectory

Then, from these measurements, the evolution of the impedance with the presence of the crack Z_f is obtained and can be written as:

$$Z_f = Z_h + \delta Z = R_f + jY_f \tag{27}$$

where the R_f and Y_f are respectively the real and imaginary parts of the indexed impedance Z_f. The remaining δZ is the evolution of the healthy impedance Z_h due to the presence of the crack. As an example, Fig.6 represents the evolution of the imaginary part of the impedance for a conductive material containing cracks with $d = 400\mu m$ depth and 2 different lengths $l = 400\mu m$ and $l = 600\mu m$ with a 2 MHz excitation frequency. The presence of the crack is clearly visible through the variation of the impedance.

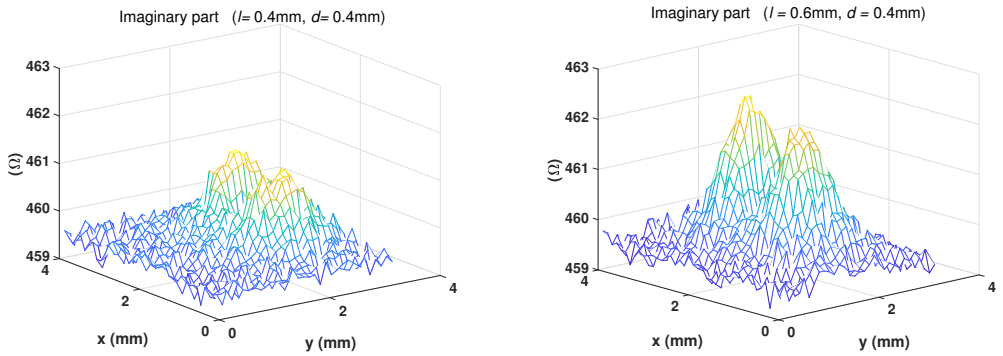


Figure 6: Imaginary part of the impedance of the ECT map for big cracks with sizes: (a) $l = 0.4mm$ and $d = 0.4mm$, (b) $l = 0.6mm$ and $d = 0.4mm$

However, in the case of smaller cracks (incipient ones), these impedance variations can be partially masked by the presence of environmental nuisance. In Fig.7 we show the evolution of the imaginary part of the impedance in the case of cracks with length or depth equal to $100\mu m$ or $200\mu m$. The evolution of the impedance value does not significantly disclose the presence of these minor cracks due to the nuisances. Therefore, we propose, in this work, to use of the Jensen-Shannon divergence to detect these incipient faults by analysing these slight impedance variation. These impedance variations will cause slight changes in the faulty probability distribution compared to the healthy reference one and are then evaluated by using the JSD.

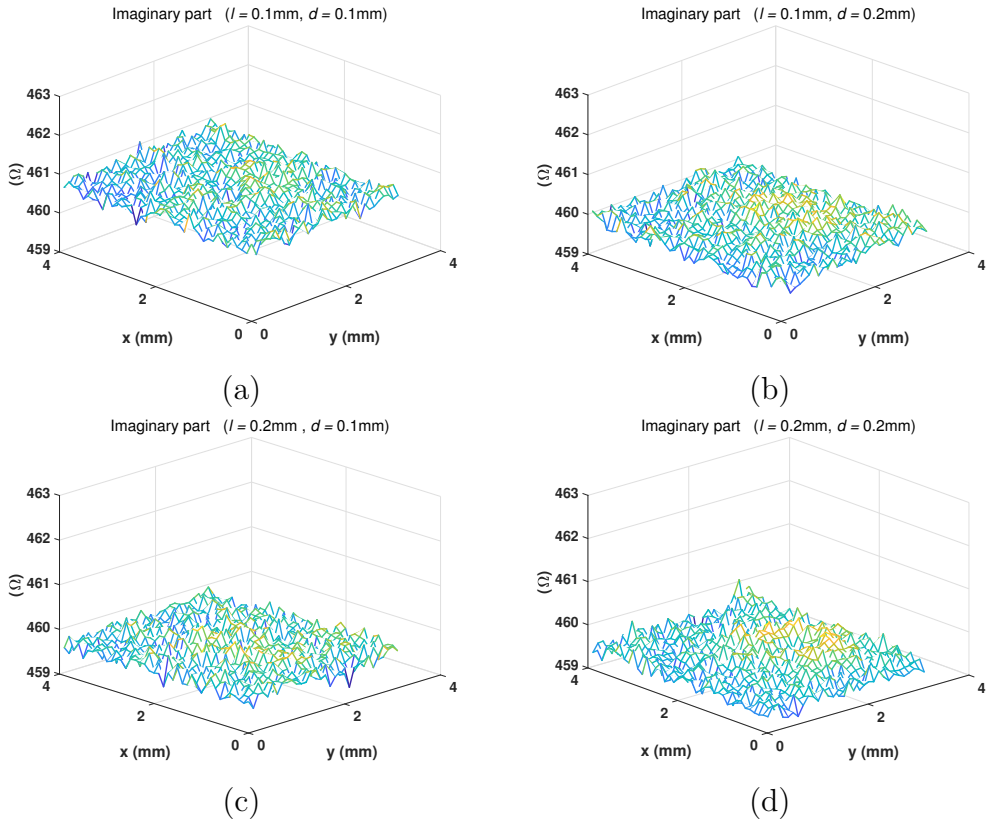


Figure 7: Imaginary part of the impedance of the ECT map for minor crack sizes: (a) $l = 0.1mm$ and $d = 0.1mm$, (b) $l = 0.1mm$ and $d = 0.2mm$, (c) $l = 0.2mm$ and $d = 0.1mm$, (d) $l = 0.2mm$ and $d = 0.2mm$

In our work, we consider the complex impedance values obtained from an ECT experimental system.

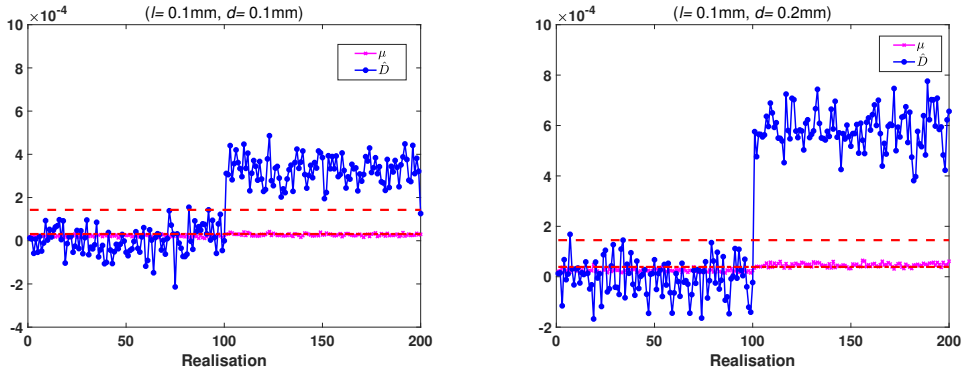
We mainly focus on incipient cracks such as their length (l) or depth (d) are equal to $100\mu m$ and $200\mu m$ in several combinations. The perturbation level will be considered using several SNR values in the range [0, 20dB]. The considered crack size for this study are then summarized in the following table:

Table 1: Considered crack sizes

Cracks	length (l)	depth (d)
C_1	$0.1mm$	$0.1mm$
C_2	$0.1mm$	$0.2mm$
C_3	$0.2mm$	$0.1mm$
C_4	$0.2mm$	$0.2mm$

To apply our diagnosis procedure, complex impedance data obtained in the healthy and the faulty conditions and considered as impedance maps with size 40×32. In our study, we focus only the imaginary part of the impedance that are known to be more sensitive to the presence of the cracks. These values of the ECT map are arranged into a single row vector (1×1280) and are then normalized. A reference map with healthy impedance values created. We enlarge each normalized ECT map (40×32) to a larger map (420×100) so that we can get the sufficient data and reduce the false alarms caused by the internal and external perturbations. For both healthy and faulty conditions 100 realisations were evaluated.

In Fig.8, we present the evolution of the JSD and the mean of the normalized impedance in healthy and faulty conditions for the crack sizes C_1 and C_2 in a 20dB SNR environment. The dashed red lines are the thresholds of the JSD and the mean calculated at 99% of their maximum values in the healthy conditions.



(a) (b)

Figure 8: JSD and mean detection capabilities for the minor cracks size: (a) C_1, (b) C_2

This result confirms that the mean fails to detect the crack with size C_1 ($P_D = 0.24$ and $P_{FA} = 0.01$). For the crack size C_2, lots of missed detections

are obtained ($P_D = 0.84$ and $P_{FA} = 0.01$). However, with the JSD, satisfying detection capabilities are obtained for both considered crack sizes ($P_D = 1$ and $P_{FA} = 0.03$).

In Fig.9 and Fig.10 we present the detection performances of JSD for the four considered incipient cracks sizes C_1, C_2, C_3, C_4 with different perturbation levels.

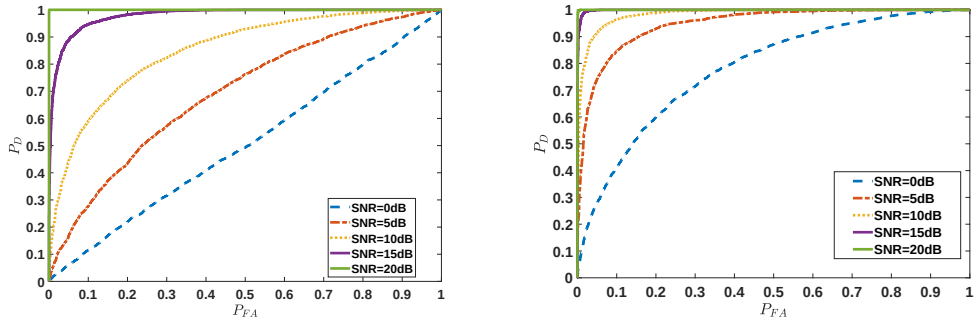


Figure 9: JSD detection performances for different perturbation level and for incipient cracks sizes: (a) C_1 and (b) C_2

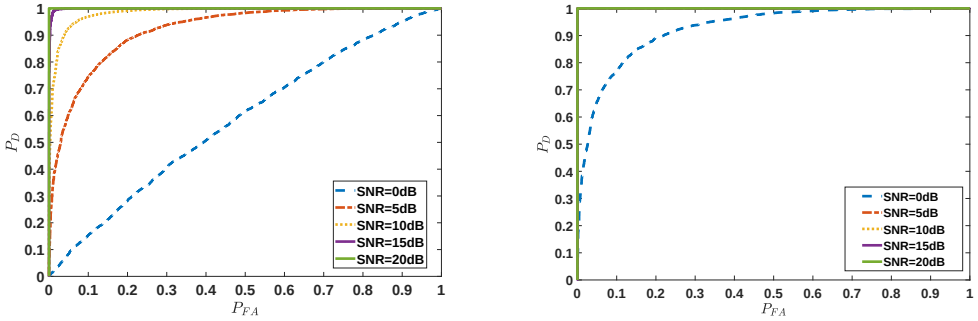


Figure 10: JSD detection performances for different perturbation levels and incipient cracks sizes: (c) C_3 and (d) C_4

The Figs clearly illustrate that the JSD has good detection performances for these incipient cracks when the SNR is larger than 10dB even when the crack size is $l = 0.1mm, d = 0.1mm$ (C_1). These detection capabilities

increase along with the SNR values. The JSD offers perfect detection performances ($P_D = 1$ with $P_{FA} = 0$) even for the smallest considered crack size (C_1) when the SNR is 20dB. This particular SNR value corresponds to the condition that the plate is smooth and the environmental perturbations are well isolated.

The perturbations that can be caused by internal or external factors (low SNR values) have a large effect on the JSD detection performances. It can be clearly noticed from Fig.9-a that the JSD is not efficient for the detection of the most incipient considered cracks size C_1 when SNR=0dB. However, when the crack size increases, the detection capabilities of JSD at 0dB increases (see Fig.9-b, Fig.10-c and Fig.10-d).

4. Incipient Fault Estimation

4.1. Fault estimation theoretical model

Based on equation (19) the fault severity g can be determined as a function of $\tilde{\lambda}_k$ and it is infinitely derivable in the neighbourhood of $g \approx 0$, The Taylor development of $\tilde{\lambda}_k$ can be written as:

$$\tilde{\lambda}_k = \lambda_k^* + \frac{\partial \tilde{\lambda}_k}{\partial g}(0)g + \frac{1}{2}\frac{\partial^2 \tilde{\lambda}_k}{\partial g^2}(0)g^2 + \cdots \tag{28}$$

In our study, the PCA is based on the covariance matrix $\mathbf{S}$ denoted as:

$$\mathbf{S} = \frac{1}{N-1}\bar{\mathbf{X}}^\intercal\bar{\mathbf{X}} = \frac{1}{N-1}\begin{bmatrix} \bar{\mathbf{x}}_1^\intercal\bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_1^\intercal\bar{\mathbf{x}}_j & \cdots & \bar{\mathbf{x}}_1^\intercal\bar{\mathbf{x}}_m \\ \vdots & & \vdots & & \vdots \\ \bar{\mathbf{x}}_j^\intercal\bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_j^\intercal\bar{\mathbf{x}}_j & \cdots & \bar{\mathbf{x}}_j^\intercal\bar{\mathbf{x}}_m \\ \vdots & & \vdots & & \vdots \\ \bar{\mathbf{x}}_m^\intercal\bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_m^\intercal\bar{\mathbf{x}}_j & \cdots & \bar{\mathbf{x}}_m^\intercal\bar{\mathbf{x}}_m \end{bmatrix} \tag{29}$$

where:

$$\bar{\mathbf{x}}_j = \mathbf{x}_j - \mu_j^{\mathbf{1}} = (\mathbf{x}_j^* - \mu_j^{*\mathbf{1}}) + (\mathrm{F}_j - g \times \frac{1}{N}\sum_{i=b}^{N} x_{ij}^{*\mathbf{1}}) + \mathbf{v}_j \tag{30}$$

where $\mathbf{1}$ is a column vector of N ones.

The first order derivative of the eigenvalue $\tilde{\lambda}_k$ can be calculated as:

$$\begin{aligned}\frac{\partial\tilde{\lambda}_k}{\partial g} &= \mathbf{p}_k^{*\intercal}\frac{\partial\mathbf{S}}{\partial g}\mathbf{p}_k^* \\ &= \frac{2}{N-1}\left(p_{jk}\sum_{r=1}^{m}p_{rk}\sum_{i=b}^{N}(x_{ir}^*-\mu_r^*)x_{ij}^* + p_{jk}^2\sum_{i=b}^{N}(x_{ij}^*-\frac{1}{N}\sum_{i=b}^{N}x_{ij}^*)^2\times g\right)\end{aligned} \tag{31}$$

where $\mathbf{p}_k^*$ is the eigenvector associated to λ_k^* in faultless and noise-free condition.

The second order derivative of $\tilde{\lambda}_k$ can be obtained as:

$$\frac{\partial^2\tilde{\lambda}_k}{\partial g^2} = \mathbf{p}_k^{*\intercal}\frac{\partial^2\mathbf{S}}{\partial g^2}\mathbf{p}_k^* = \frac{2}{N-1}\left(p_{jk}^2\sum_{i=b}^{N}(x_{ij}^*-\frac{1}{N}\sum_{i=b}^{N}x_{ij}^*)^2\right) \tag{32}$$

if we denote that:

$$a_1 = p_{jk}\sum_{r=1}^{m}p_{rk}\left(\sum_{i=b}^{N-1}(x_{ir}^*-\mu_r^*)x_{ij}^*\right) \tag{33}$$

$$a_2 = 3p_{jk}^2\sum_{i=b}^{N}\left(x_{ij}^*-\frac{1}{N}\sum_{i=b}^{N}x_{ij}^*\right)^2 \tag{34}$$

Choosing the first two derivative order of the equation (28) we obtain the following equation:

$$\Delta\lambda_k = \frac{2}{N-1}a_1\times g + \frac{1}{N-1}a_2\times g^2 \tag{35}$$

Then, the theoretical fault estimation expression can be obtained by finding the positive solution of (35):

$$\hat{g} = \frac{-a_1+\sqrt{a_1^2+(N-1)a_2\Delta\lambda_k}}{a_2} \tag{36}$$

In order to obtain the fault amplitude estimation equation, we must calculate the equation of $\Delta\lambda_k$ corresponding to the JSD values. As equation (21) can be seen as a function of $\Delta\lambda_k$ and it is infinitely derivable in the neighborhood of zero, the Taylor development of D_{JS} can be given following (37).

$$D_{JS}(\Delta\lambda_k) = D_{JS}(0)+\frac{\partial D_{JS}(\Delta\lambda_k)}{\partial\Delta\lambda_k}(0)\Delta\lambda_k+\frac{1}{2}\frac{\partial^2 D_{JS}(\Delta\lambda_k)}{\partial\Delta\lambda_k^2}(0)\Delta\lambda_k^2+\cdots \quad (37)$$

We obtain its first order derivative (38) and second order derivative (39) based on (21).

$$a_3 = \frac{\partial D_{JS}(\Delta\lambda_k)}{\partial\Delta\lambda_k} = \frac{1}{4}\left[\frac{2}{2\lambda_k^* + \Delta\lambda_k + 2\sigma_v^2} - \frac{1}{\lambda_k^* + \Delta\lambda_k + \sigma_v^2}\right] \quad (38)$$

$$a_4 = \frac{\partial^2 D_{JS}(\Delta\lambda_k)}{\partial\Delta\lambda_k{}^2} = \frac{1}{4}\left[\frac{1}{(\lambda_k^* + \Delta\lambda_k + \sigma_v^2)^2} - \frac{2}{(2\lambda_k^* + \Delta\lambda_k + 2\sigma_v^2)^2}\right] \quad (39)$$

Choosing the first two order derivatives of Taylor equation to establish a quadratic equation (40) from which the resolution can derive the approximated value of $\Delta\lambda_k$.

$$D_{JS}(\Delta\lambda_k) = D_{JS}(0) + a_3(0)\Delta\lambda_k + \frac{1}{2}a_4(0)\Delta\lambda_k^2 \quad (40)$$

In healthy conditions, the variance change $\Delta\lambda_k$ is 0, and $a_3(0)$ and $a_4(0)$ can be simplified as in (41). Then we find the solution given in (42).

$$a_3(0) = 0 \qquad a_4(0) = \frac{1}{4}[\frac{1}{(\lambda_k^* + \sigma_v^2)^2} - \frac{2}{(2\lambda_k^* + 2\sigma_v^2)^2}] \quad (41)$$

$$\Delta\lambda_k = \sqrt{\frac{2\hat{D}_{JS}}{a_4(0)}} \quad (42)$$

Finally, the approximated theoretical estimation of the fault severity $\hat{g}$ can be obtained by combining (36) and (42) such as:

$$\hat{g} = \frac{-a_1 + \sqrt{a_1^2 + (N-1)a_2(\sqrt{\frac{2\hat{D}_{JS}}{a_4(0)}})}}{a_2} \quad (43)$$

where $\hat{D}_{JS}$ in (43) is calculated using Monte Carlo Simulation. Based on this equation (43), the fault severity can be estimated using the JSD value.

4.2. Incipient Fault estimation performances results and discussion

To verify the validity of the fault estimation model obtained in (43) for incipient fault, we plot in Fig.11 the estimated fault severity ($\hat{g}$) versus the real one (g) for different SNR values. For this work we have considered the same AR Model that has been used in the previous section 3.2.

The results indicate that the fault estimation obtained with this model leads to a little overestimation of the fault severity. This can be considered as a safety margin in a sensible fault diagnosis context.

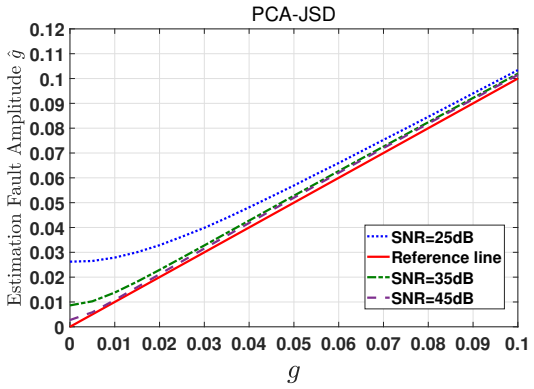


Figure 11: Fault estimation results at different noise levels

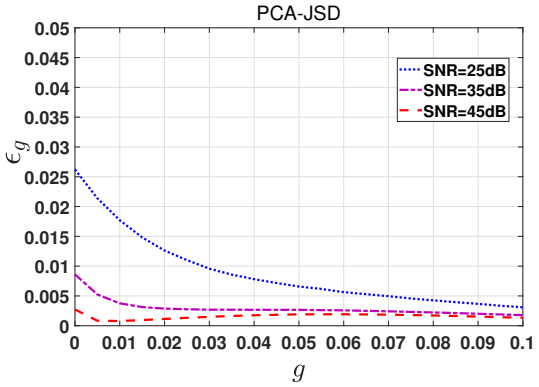


Figure 12: Estimation relative error at different noise levels

The relative error ϵ_g is calculated as $\epsilon_g = \frac{\hat{g}-g}{1+g}$ for quantifying the overestimation level. The results for different SNR are presented in Fig.12. The

relative error decreases either while the fault severity or the SNR increases. For SNR = 25dB, the maximum error is 2.75%.

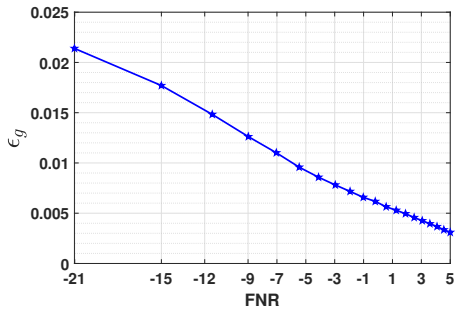


Figure 13: Estimation relative error with SNR=25dB for different FNR

The relative error results for different FNR (with SNR set at 25dB) is displayed in Fig.13. The result highlights that the maximum error is 2.15% when FNR is −21dB. Such a value is sufficiently low to be acceptable considering the incipient fault severity diagnosis context.

4.3. Application to nondestructive incipient cracks severity estimation

In order to validate the proposed estimation process, we have considered the evaluation of cracks based on experimental ECT data. The considered ECT cracks considered here are the same as those described in section 3. The fault estimation model is focused on the crack sizes given in Table 2 where the crack area increases along with the length or depth values. In this work only incipient cracks evaluation are considered. Then no cracks with an area larger than $0.24mm^2$ are evaluated here. These restrictions imply a more tedious evaluation. For the estimation model, considering such constraint, we have used the Taylor equation to be able to approximately derive some part of the equations. This leads to the assumption that the estimated value should be close to zero and not too far. While this assumption is not respected, the estimation accuracy is decreased.

For each incipient crack size, we compute the FNR value. For this latter, the noise level is related to the internal and external sources. It is obtained considering the variance of the impedance signal in healthy conditions. Assuming that the fault is independent from the noise, the fault level is considered as the main evolution between healthy and faulty signals. As shown

in Table 2, FNR for the smallest and biggest crack sizes are respectively -29.3dB and 0.21dB.

Table 2: Crack sizes area and corresponding experimental FNR

Area (mm^2)	0.01	0.02	0.04	0.08	0.16	0.24
length (l), depth (d)	$0.1, 0.1$	$0.1, 0.2$	$0.2, 0.2$	$0.4, 0.2$	$0.4, 0.4$	$0.6, 0.4$
Experimental FNR(dB)	-29.3	-20.1	-16.8	-9.63	-4.5	0.21

For our study, we consider the reference healthy data as the ECT matrix $\mathbf{X}_h$ denoted as $\mathbf{X}_h = (\mathbf{z}_1^h, \mathbf{z}_2^h, \mathbf{z}_3^h, \mathbf{z}_4^h)$, where $\mathbf{z}^h$ is the enlarged imaginary part of the impedance signals acquired from the edge of the normalized map. The faulty ECT matrix $\mathbf{X}_f$ denoted as $\mathbf{X}_f = (\mathbf{z}_1^f, \mathbf{z}_2^f, \mathbf{z}_3^f, \mathbf{z}_4^f)$, where $\mathbf{z}^f$ is the enlarged imaginary impedance signals acquired from the faulty normalized map. The number of samples is $N = 4.2 * 10^4$.

The PCA is applied to $\mathbf{X}_h$ and $\mathbf{X}_f$. Then the reference distribution and the faulty distribution of the first principal component scores are calculated. Fig.14 presents the probability distribution for the first principal component score $\mathbf{t}_1$ corresponding to the crack size C_1. Twenty acquisitions of $\mathbf{z}^f$ are done for each crack size. The Jensen-Shannon divergence values between a faulty distribution and the reference one are computed according to equation (12) using Monte carlo simulation. The estimated fault severities impedance values are obtained based on equation (43). In Fig.15 we plot the average estimation of the fault severity according to the crack size measured by FNR value.

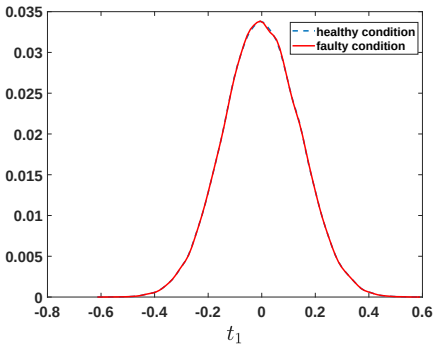


Figure 14: Pdfs of $\mathbf{t}_1$ for the healthy and faulty component scores

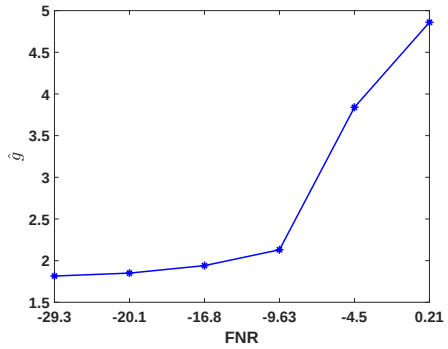


Figure 15: Fault severity estimation versus the FNR(dB)

Fig.15 clearly shows that the estimated fault severity grows monotonically with the FNR. This result validates the effectiveness of the proposed fault estimation model.

5. Conclusion

In this paper, we have proposed an incipient fault diagnosis scheme based on the Jensen-Shannon Divergence for fault detection and estimation. For both cases, a theoretical model is proposed and widely show the dependence of the fault with the JSD and the noise level.

The effectiveness of the detection is first validated by using simulated data from an AR system. The detection performances of JSD for incipient fault are compared to the Hotelling's test T^2 and the SPE ones. Using ECT experimental data to detect incipient cracks in a conductive plate, JSD good performances have been obtained in terms of P_D and P_{FA} even with SNR as low as 20dB. The proposed approach allows to evaluate such incipient cracks that was not able to be detected in the literature with such noisy ECT data.

Concerning the fault estimation, we prove, for Gaussian distributed data, the efficiency and the accuracy of the proposal. For low noise conditions, the estimation accuracy is good, the maximum relative error is 2.75% for SNR=25dB and low fault severities (the maximum relative error is 2.15% for $FNR = -21$dB). With our proposal, the fault severity is slightly overestimated particularly in high noise condition levels. This slight overestimation is a valuable safety margin in health monitoring process. Finally, the application for fault severity estimation on the ECT data proves the efficiency of

the derived model by providing an accurate estimation of the fault amplitude versus the FNR.

Acknowledgement

This research was partially supported by the iCODE Institute, research project of the IDEX Paris-Saclay, and by the Hadamard Mathematics LabEx (LMH) through the grant number ANR-11-LABX-0056-LMH in the "Programme des Investissements d'Avenir".

The authors would like to thank China Scholarship Council for funding and also Professor Y. Le Bihan for providing the experimental ECT data.

References

[1] M. A. Demetriou, M. M. Polycarpou, Incipient fault diagnosis of dynamical systems using online approximators, IEEE Transactions on Automatic Control 43 (11) (1998) 1612–1617.

[2] J. Shang, M. Chen, H. Ji, D. Zhou, Recursive transformed component statistical analysis for incipient fault detection, Automatica 80 (2017) 313–327.

[3] Z. He, Y. A. Shardt, D. Wang, B. Hou, H. Zhou, J. Wang, An incipient fault detection approach via detrending and denoising, Control Engineering Practice 74 (2018) 1–12.

[4] Q. Wang, C. Wang, Incipient fault detection of nonlinear dynamical systems via deterministic learning, Neurocomputing 313 (2018) 125–134.

[5] H. Ji, X. He, J. Shang, D. Zhou, Incipient fault detection with smoothing techniques in statistical process monitoring, Control Engineering Practice 62 (2017) 11–21.

[6] K. E. S. Pilario, Y. Cao, Canonical variate dissimilarity analysis for process incipient fault detection, IEEE Transactions on Industrial Informatics 14 (12) (2018) 5308–5315.

[7] S. J. Qin, Survey on data-driven industrial process monitoring and diagnosis, Annual Reviews in Control 36 (2) (2012) 220 – 234.

[8] F. Harrou, M. N. Nounou, H. N. Nounou, M. Madakyaru, Statistical fault detection using PCA-based GLR hypothesis testing, Journal of loss prevention in the process industries 26 (1) (2013) 129–139.

[9] V. Venkatasubramanian, R. Rengaswamy, s. Kavuri, K. Yin, A review of process fault detection and diagnosis part I: Quantitative model-based methods, Elsevier Journal On Computer and Chemical Engineering 27 (2003) 293–311.

[10] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, K. Yin, A review of process fault detection and diagnosis part III: Process history based methods, Elsevier Journal On Computer and Chemical Engineering 27 (2003) 327–346.

[11] C. Delpha, D. Diallo, H. Al Samrout, N. Moubayed, Incipient fault detection and diagnosis in a three-phase electrical system using statistical signal processing, in: IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2017, pp. 3828–3833.

[12] S. Yin, X. Ding, X. Xie, H. Luo, A review on basic data-driven approach for industrial process monitoring, IEEE Transactions on Industrial Electronics 61 (11) (2014) 6418–6428.

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

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

[15] M. Basseville, Distances measures for signal processing and pattern recognition, Elsevier Signal Processing 18 (4) (1989) 349–369.

[16] M. Basseville, Divergence measures for statistical data processing–an annotated bibliography, Signal Processing 93 (4) (2013) 621–633.

[17] J. Briët, P. Harremoës, Properties of classical and quantum Jensen-Shannon divergence, Physical review A 79 (5) (2009) 052311.

[18] X. Huang, S. Z. Li, Y. Wang, Jensen-shannon boosting learning for object recognition, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, IEEE, 2005, pp. 144–149.

[19] W. Yang, H. Song, X. Huang, X. Xu, 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 7 (8) (2014) 3318–3327.

[20] T. M. Osán, D. G. Bussandri, P. W. Lamberti, Monoparametric family of metrics derived from classical Jensen–Shannon divergence, Physica A: Statistical Mechanics and its Applications 495 (2018) 336–344.

[21] A. Mehri, M. Jamaati, H. Mehri, Word ranking in a single document by Jensen–Shannon divergence, Physics Letters A 379 (28-29) (2015) 1627–1632.

[22] S. Molladavoudi, H. Zainuddin, K. T. Chan, Jensen–Shannon divergence and non-linear quantum dynamics, Physics Letters A 376 (26-27) (2012) 1955–1961.

[23] Y. Tharrault, G. Mourot, J. Ragot, D. Maquin, Fault detection and isolation with robust principal component analysis, International Journal of Applied Mathematics and Computer Science 18 (4) (2008) 429–442.

[24] H. Wang, Z. Song, P. Li, Fault detection behavior and performance analysis of principal component analysis based process monitoring methods, Industrial & Engineering Chemistry Research 41 (10) (2002) 2455–2464.

[25] J. Harmouche, C. Delpha, D. Diallo, Incipient fault detection and diagnosis based on Kullback–Leibler divergence using Principal Component Analysis: Part I, Signal Processing 94 (2014) 278–287.

[26] J. Harmouche, C. Delpha, D. Diallo, Incipient fault detection and diagnosis based on Kullback–Leibler divergence using Principal Component Analysis: Part II, Signal Processing 109 (2015) 334–344.

[27] I. Joliffe, B. Morgan, Principal component analysis and exploratory factor analysis, Statistical methods in medical research 1 (1) (1992) 69–95.

[28] R. L. Mason, J. C. Young, Implementing multivariate statistical process control using hotelling's T^2 statistics, Quality Progress 34 (4) (2001) 71.

[29] J. P. George, Z. Chen, P. Shaw, Fault detection of drinking water treatment process using PCA and hotellings T^2 chart, World Academy of Science, Engineering and Technology 50 (2009) 970–975.

[30] Y. Du, D. Du, Fault detection and diagnosis using empirical mode decomposition based principal component analysis, Computers & Chemical Engineering 115 (2018) 1–21.

[31] R. Hamia, C. Cordier, C. Dolabdjian, Eddy-current non-destructive testing system for the determination of crack orientation, NDT & E International 61 (2014) 24–28.

[32] G. Y. Tian, A. Sophian, Defect classification using a new feature for pulsed eddy current sensors, Ndt & E International 38 (1) (2005) 77–82.

[33] M. R. Bato, A. Hor, A. Rautureau, C. Bes, Impact of human and environmental factors on the probability of detection during NDT control by eddy currents, Measurement 133 (2019) 222–232.

[34] P. Zhu, Y. Cheng, P. Banerjee, A. Tamburrino, Y. Deng, A novel machine learning model for eddy current testing with uncertainty, NDT & E International 101 (2019) 104–112.

[35] A. Sophian, G. Y. Tian, D. Taylor, J. Rudlin, A feature extraction technique based on principal component analysis for pulsed eddy current ndt, NDT & e International 36 (1) (2003) 37–41.

[36] Y. He, M. Pan, F. Luo, D. Chen, X. Hu, Support vector machine and optimised feature extraction in integrated eddy current instrument, Measurement 46 (1) (2013) 764–774.

[37] X. Chen, D. Hou, L. Zhao, P. Huang, G. Zhang, Study on defect classification in multi-layer structures based on fisher linear discriminate analysis by using pulsed eddy current technique, NDT & E International 67 (2014) 46–54.

[38] J. Harmouche, C. Delpha, D. Diallo, Y. Le Bihan, Statistical approach for nondestructive incipient crack detection and characterization using

Kullback-Leibler divergence, IEEE Transactions on Reliability 65 (3) (2016) 1360–1368.

[39] N. Taki, W. B. Hassen, N. Ravot, C. Delpha, D. Diallo, Frequency selection for reflectometry-based soft fault detection using principal component analysis, in: 2019 Prognostics and System Health Management Conference (PHM-Paris), IEEE, 2019, pp. 273–278.

[40] J. L. Horn, A rationale and test for the number of factors in factor analysis, Psychometrika 30 (2) (1965) 179–185.

[41] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. on Information Theory 37 (1) (1991) 145–151. doi:10.1109/18.61115.

[42] T. Cover, J. Thomas, Elements of Information Theory, Wiley and sons: New Jersey (USA), 2005.

[43] J. Harmouche, C. Delpha, D. Diallo, Incipient fault amplitude estimation using KL divergence with a probabilistic approach, Signal Processing 120 (2016) 1–7.

[44] I. Eisenberger, Genesis of bimodal distributions, Technometrics 6 (4) (1964) 357–363.