Détection et diagnostic de défauts naissants en utilisant la divergence de Kullback-Leibler : De la théorie aux applications

Jinane Harmouche

ÉCOLE DOCTORALE : STITS

Laboratoire des Signaux et Systèmes (L2S, UMR 8506)
Laboratoire de Génie Electrique de Paris (LGEP, UMR 8507)

DISCIPLINE *PHYSIQUE*
Spécialité
Traitement de signal

THÈSE DE DOCTORAT

soutenue le **20/11/2014**

par

Jinane HARMOUCHE

Statistical Incipient Fault Detection and Diagnosis with Kullback-Leibler Divergence : From Theory to Applications

Directeur de thèse :	Claude DELPHA	Maître de Conférences HDR (L2S, Univ. Paris-Sud)
Co-directeur de thèse :	Demba DIALLO	Professeur des univ. (LGEP, Univ. Paris-Sud)
Composition du jury :		
Président du jury :	Nadine MARTIN	Directeur de Recherche (Gipsa-Lab, Grenoble INP)
Rapporteurs :	François AUGER	Professeur des univ. (IREENA, Univ. de Nantes)
	Guy CLERC	Professeur des univ. (Ampère, Univ. Claude Bernard)
Examinateurs :	Mohamed BENBOUZID	Professeur des univ. (LBMS, Univ. de Brest)
	Stéphane FONT	Professeur (E3S, Supélec)

Résumé

Les travaux de cette thèse portent sur la détection et le diagnostic des défauts naissants dans les systèmes d'ingénierie et industriels, par des approches statistiques non-paramétriques. Un défaut naissant est censé provoquer comme tout défaut un changement anormal dans les mesures des variables du système. Ce changement est cependant imperceptible mais aussi imprévisible dû à l'important rapport signal-sur-défaut, et le faible rapport défaut-sur-bruit caractérisant le défaut naissant. La détection et l'identification d'un changement général nécessite une approche globale qui prend en compte la totalité de la signature des défauts. En plus, le défaut naissant a une faible signature et sa détection fait appel aux indicateurs ayant une haute sensibilité aux petites distorsions. Dans ce cadre, la divergence de Kullback-Leibler est proposée comme indicateur général de défauts, sensible aux petites variations anormales cachées dans les variations du bruit. Une approche d'analyse spectrale globale est également proposée pour le diagnostic de défauts ayant une signature fréquentielle. L'application de l'approche statistique globale est illustrée sur deux études différentes. La première concerne la détection et la caractérisation, par courants de Foucault, des fissures dans les structures conductrices. L'approche basée sur la divergence est appliquée à des mesures expérimentales d'impédance d'une plaque conductrice présentant plusieurs fissures de petites tailles. La deuxième application concerne le diagnostic des défauts de roulements dans les machines électriques tournantes. L'approche basée sur une analyse globale du spectre est appliquée sur des signaux vibratoires générés, à plusieurs niveaux de charge, pour différents types et tailles de défauts de roulements. En outre, ce travail traite le problème d'estimation de l'amplitude des défauts naissants. Une analyse théorique menée dans le cadre d'une modélisation par analyse en composantes principales, conduit à un modèle analytique de la divergence ne dépendant que des paramètres du défaut. Un estimateur de l'amplitude du défaut est ainsi obtenu. Les performances de détection (au travers des probabilités de non détection et de fausse alarme) et la précision de l'estimation des défauts naissants sont évaluées sur des exemples numériques.

Abstract

This phD dissertation deals with the detection and diagnosis of incipient faults in engineering and industrial systems by non-parametric statistical approaches. An incipient fault is supposed to provoke an abnormal change in the measurements of the system variables. However, this change is imperceptible and also unpredictable due to the large signal-to-fault ratio and the low fault-to-noise ratio characterising the incipient fault. The detection and identification of a global change require a 'global' approach that takes into account the total faults signature. In addition, incipient faults have a weak signature that requires detection indicators that have high sensitivity to small distortions. In this context, the Kullback-Leibler divergence is considered to be a 'global' fault indicator, which is recommended sensitive to abnormal small variations hidden in noise. A 'global' spectral analysis approach is also proposed for the diagnosis of faults with a frequency signature. The 'global' statistical approach is proved on two application studies. The first one concerns the detection and characterization of minor cracks in conductive structures. The approach based on the divergence is applied to experimental impedance signals from the eddy-current testing of a conductive plate that has a plurality of small-sized cracks. The second application concerns the diagnosis of bearing faults in electrical rotating machines. The 'global' spectral analysis approach is applied to experimental vibration signals generated by different types and sizes of bearing faults, for different operating points. In addition, the fault estimation problem is addressed in this work. A theoretical study is conducted to obtain an analytical model of the KL divergence, from which an estimate of the amplitude of the incipient fault is derived. The fault detection performances (false alarm and missed detection probabilities) of the divergence and the fault estimation accuracy are evaluated on numerical models through intensive simulations.

List of publications

This PhD work has led to 3 journal papers, 5 international conference papers and 2 national conference papers.

International journal papers

J3 **J. Harmouche**, C. Delpha, D. Diallo, "Improved Fault diagnosis of ball bearings based on the global spectrum of vibration signals", IEEE Transactions on Energy Conversion, Accepted for publication, DOI : 10.1109/TEC.2014.2341620

J2 **J. Harmouche**, C. Delpha, and D. Diallo, "Incipient Fault Detection and Diagnosis Based on Kullback-Leibler Divergence Using Principal Component Analysis: Part II", Signal Processing, Accepted for publication, DOI : 10.1016/j.sigpro.2014.06.023.

J1 **J. Harmouche**, C. Delpha, and D. Diallo, "Incipient Fault Detection and Diagnosis Based on Kullback-Leibler Divergence Using Principal Component Analysis: Part I", Signal Processing, vol. 94 p.278-287 (2014).

International conference papers

CI5 **J. Harmouche**, C. Delpha, and D. Diallo, "Linear Discriminant Analysis for the Discrimination of Faults in Bearing Balls by using Spectral Features" - Conference: IEEE ICGE 2014 (Sfax, TN, 2014-03-25) - Proceedings of the IEEE International Conference on Green Energy, (2014) (Conference Best paper Award).

CI4 A. Youssef, **J. Harmouche**, C. Delpha and D. Diallo, "Capability Evaluation of Incipient Fault Detection in Noisy Environment: A Theoretical Kullback-Leibler Divergence-Based Approach for Diagnosis" - Conference: IECON 2013 (Vienne, AT, 2013-11-10), Proceedings of the 39th Annual Conference of the IEEE Industrial Electronics Society, (2013)

CI3 **J. Harmouche**, C. Delpha and D. Diallo, "A Theoretical Approach for Incipient Fault Severity Assessment Using the Kullback-Leibler Divergence" - Conference: EUSIPCO 2013 (Marrakech, MA, 2013-09-09), Proceedings of the IEEE European Signal Processing Conference, (2013)

CI2 **J. Harmouche**, C. Delpha and D. Diallo, "A Global Approach for Classification of Bearing Conditions Using Spectral Features" - Conference: IECON 2013 (Vienne, AT, 2013-11-10) - Proceedings of the 39th Annual Conference of the IEEE Industrial Electronics Society, (2013)

CI1 **J. Harmouche**, C. Delpha and D. Diallo, "Faults diagnosis and detection using Principal Component Analysis and Kullback-Leibler divergence" - Conference: IEEE IECON 2012 (Montréal, CA, 2012-10-25), 38th Annual Conference of the IEEE Industrial Electronics Society, (2012) (Session Best paper Award)

National (french) conference papers

CN2 **J. Harmouche**, D. Diallo, and C. Delpha, "Discrimination des défauts de roulements par une analyse spectrale globale" - Conférence: SGE 2014 (Cachan, FR, 2014-07-08) - Actes du Symposium de Génie Electrique, (2014)

CN1 **J. Harmouche**, C. Delpha and D. Diallo, "Estimation théorique de la sévérité d'un défaut par un modèle analytique de la DKL" - Conférence: GRETSI 2013 (Brest, FR, 2013-09-03) - Actes du 24ème Colloque GRETSI, (2013)

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

ANN	Artificial Neural Networks
AR	Autoregressive
ARMA	Autoregressive Moving Average
BF	Ball fault
ECT	Eddy Currents Testing
EDM	Electro discharge-machining
EVD	Eigenvalue decomposition
FDD	Fault Detection and Diagnosis
FFT	Fast Fourrier Transform
FNR	Fault-to-Noise Ratio
FTC	Fault Tolerant Control
GLR	generalized likelihood ratio
ICA	Independent Component Analysis
IRF	Inner race fault
KL	Kullback-Leibler
LDA	Linear Discriminant Analysis
LV	Latent Variables
MCUSUM	Multivariate Cumulative Sum
MEWMA	Multivariate Exponentially Weighted Moving Average
NDI	Non-Destructive Inspection
NDT	Non-Destructive Testing
ORF	Outer race fault
PCA	Principal Component Analysis
pdf	probability distribution function
P_{FA}	Probability of false alam
PLS	Projection to Latent Structure
P_{MD}	Probability of missed detection
SFR	Signal-to-Fault Ratio

SNR	Signal-to-Noise Ratio
SPE	Squared Prediction Error
SVM	Support Vector Machines
$\mathbf{T^2}$	Hotelling $\mathbf{T^2}$

Symbols

a	fault amplitude
b	fault start
C_i	class i
d_c	crack depth
d_b	ball diameter of bearing
D_b	pitch diameter of bearing
$\hat{D}$	approximated KL divergence
D_{An}	analytical KL divergence
$\hat{D}_{t_k}$	KL divergence approximated on the kth principal component score
E_r	relative estimation error
f_c	bearing cage rotational frequency
f_{exc}	excitation frequency
f_{fa}	characteristic frequency
$\boldsymbol{f_i}$	ith specific frequency/spectral feature
f_k	pdf of the kth principal component score
f_k^{rf}	reference pdf of the kth principal component score
f_r	shaft rotational frequency
I_m	identity matrix
l	principal subspace dimension
l_c	crack length
m	dimension of the process data/features vector
N	data size
n_b	number of bearing balls
P	matrix of eigenvectors
$\boldsymbol{p_k}$	kth eigenvector
p_{jk}	jth coefficient of the kth eigenvector
R	realisations
$\boldsymbol{R_0}, \boldsymbol{R_c}, \boldsymbol{R_d}$	real impedance part
S	covariance/correlation matrix
S_i	covariance matrix of classe C_i
S_B	between classes covariance matrix
S_w	within classes covariance matrix
T	principal component scores matrix

$\boldsymbol{t_k}$	kth principal component score
t_{ik}	ith observation of the kth principal component score
$t(i)$	ith observation of principal component scores
X	data matrix
$\boldsymbol{x_j}$	jth variable
x_{ij}	ith observation of the jth variable
$x(i)$	ith observation of variables
$\boldsymbol{X_0}, \boldsymbol{X_c}, \boldsymbol{X_d}$	imaginary impedance part
$\boldsymbol{Z_0}, \boldsymbol{Z_c}, \boldsymbol{Z_d}$	complex impedance
λ_k	kth eigenvalue
ϵ_{safe}	threshold for the KL divergence
σ_v^2	noise variance
σ_f^2	fault variance
μ	mean
ϕ	bearing contact angle
* mark	fault-free, noiseless observations
$^-$mark	normalised observations
' mark	transpose operator

General introduction

Context of the study

The last three decades have shown an increased demand for improving the economy and safety of processes. Health monitoring of processes has been widely developed with studies of fault detection and diagnosis. Initially seen as an application area of signal processing, filter, and control theory, the fault detection and diagnosis field continues to be the focus of most researches. In a wide variety of industrial and on-board applications, the detection and diagnosis of faults are considered essential to ensure high performance level of the plant operation, to reduce economic losses and to enhance the security of a plant operating in a controllable region. Aerospace systems, automotive applications, power systems and industrial production systems are popular examples of these applications.

A plant or system consists roughly of three main subsystems, actuators, main process and sensors. Either of the system's components is vulnerable to faults due to its interaction with the environment, ageing, manufacturing defects, harsh operation conditions, misuse, human errors, etc. A 'fault' denotes a tolerable malfunction which potentially may develop into a failure if it has not been detected and isolated early enough so that appropriate corrective actions can be taken [1]. **The fault detection** consists of deciding whether the system operation has deviated from the normal standard conditions due to an undesired behaviour of one of the system components. **The fault isolation** follows the fault detection task and consists of determining the location of the fault, or which sensor, actuator or process component has become faulty. In some applications, it is also important to determine the time occurrence of the fault and to assess its severity. This task is defined by **fault identification**. Monitoring a physical system requires continuously observing its behaviour to detect, isolate and identify faults. A procedure including the three mentioned tasks is called a fault detection and diagnosis (FDD) system.

There are extensive theoretical and practical studies that investigate various approaches for FDD. Whatever the approach, the FDD system primarily relies

upon two elements: availability of process measurements and prior knowledge of the potential faults. Process measurements are a basic representation of the process behaviour. They however carry information about the faults, usually called the signature or symptoms of faults. The objective of a FDD system is to extract the fault-related information from the process measurements by conducting appropriate transformations on the process data. These transformations lead to some useful features that help assessing the operating health state and reflecting the potential faults. The features can be in the form of changes in spectral properties, in statistical parameters, in some functional relationships among the process variables (states, inputs and outputs), qualitative rules, etc. They can be obtained using either quantitative physical models (input/output models, state/space models, first-principles models, frequency response models, etc.), qualitative models (expert systems based on if-then-else rules, fault trees, cause-effect models, etc.) or data-driven implicit models (process history based) [2, 3].

What is the suitable transformation/modelling to be done? In fact, the FDD system design and specifically the structure of the model to be built are imposed by the dedicated application. The application usually specifies the type and amount of knowledge provided for its users and the desirable FDD characteristics. The FDD system uses this *a priori* process knowledge to transform the process data into features that achieve the desired FDD characteristics or properties. For example, to develop an accurate quantitative model for the system under consideration a sufficient amount of information about the process physics including all forms of interactions with the environment should be available. Aerospace and electromechanical engineering processes are application examples. However, when dealing with complex, large scaled, time varying or nonlinear processes, a comprehensive model becomes unfeasible and expensive to develop and to be effective. This is the case of process industries and chemical plants for instance. Introducing simplification hypothesis or assumptions to models limits their generalisation capabilities to certain chosen operation modes and induces modelling errors that reduce greatly the effectiveness of the FDD system. These problems are likewise encountered in some emerging technologies, like the fuel cells that are embedded in transport applications. If the information available about the operation modes is qualitative rather than quantitative, the model to be developed will be also qualitative. For the qualitative model to be complete and effective, abundance of experience on the process under normal and faulty operating conditions is required. However, if a little amount of information on the process mechanisms is available and measurement signals are the only valuable resource for process monitoring, data-driven implicit models would be the best candidates.

Data-driven models are implicit empirical models derived from analysis of available data. Their derivation requires a minimal *a priori* knowledge about process physics, but a significant amount of historical process data which may contain faults and their symptoms. It is mainly based on computational intelligence and machine-learning methods, which are referred to as data-driven methods. These include multivariate statistical projection methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN) and Support Vector Machines (SVM). PCA and Projection to Latent Structure (PLS) are among the data-driven methods, which have gained a remarkable acceptance in industry for statistical monitoring and control of multivariate processes [4]. They model the correlation structure existing among the process variables. They are especially useful when the number of variables is large enough so that their variation is likely due to a small number of underlying relevant variables.

The multivariate statistical methods attempt to analyse high dimensional data in order to capture the underlying structure formed with some Latent Variables (LV) that reveal some statistical characteristics. The latent variables are generally decomposed into dominant and residual ones, therefore reducing the dimension of the original variable space. The process statistical monitoring based on these methods relies on statistics that detect deviation of a new observation from the LV control region. The Hotelling $\mathbf{T^2}$ and the Squared Prediction Error (SPE) are typical distance-based statistics commonly applied for the on-line monitoring. The SPE, computed over residual LV, is sensitive to atypical observations having abnormal high magnitude. The $\mathbf{T^2}$, computed over dominant LV, is less sensitive to abnormal observations, but is able to detect moderate to large shifts in the mean or the variance of variables. The multivariate control charts [5, 6], such as the Multivariate Exponentially Weighted Moving Average (MEWMA) and the Multivariate Cumulative Sum (MCUSUM) can be used to detect mean shifts, and they also show a higher sensitivity to small shifts than $\mathbf{T^2}$.

It is worth mentioning that the statistical analysis of process data may be performed in any informative domain, including the time and/or the frequency domains. At first, time-series data are transformed into appropriate quantities like frequencies, amplitude, time-scale features and phase characteristics, using advanced signal processing tools [7]. This can lead to univariate or multivariate features. Then, the multivariate statistical projection methods can be applied to the potential multivariate features vector in order to depict the underlying relevant variability among redundant features. This is called a feature extraction task. The extracted features are generally provided to a classification system for on-line monitoring. The

choice of features is very influential on the performance and complexity of the FDD system because depending on their relevance, the classification algorithm can be either very complex or quite simple. Having an important differentiation power, informative features, also called high-level features, are always in need, and they usually involve sophisticated signal processing techniques such as time-scale analysis, time-frequency analysis and advanced denoising. This is particularly the case with condition monitoring and fault diagnosis in rotating machinery [8, 9]. Electrical machine fault diagnosis relies primarily on the analysis of available signals into different representation domains in order to extract useful features for classification. Vibration analysis, despite being expensive, is asserted to provide the most effective features to diagnose mechanical problems [10]. Motor Current Signature Analysis (MCSA) is powerful at detecting and diagnosing electrical faults [11]. However, since a rotating machine is an interconnected electromechanical system, mechanical and electrical signals are often analysed simultaneously to accurately assess the machine health and identify the fault's origin.

Objectives

Statistically formulating the process, the probability distributions of observations related to the process variables summarize the statistical characteristics of the process. When a process operates under normal mode, its variables will have probability distributions corresponding to the fault-free operating conditions. Most statistics for fault detection are designed to detect shifts in the distribution parameters, such as mean and variance, of the process variables.

Real faults, at early stage of development, can nevertheless affect the probability distributions in an unpredictable (random) manner causing slight changes and distortions along the distributions rather than a certain change in one of its parameters. This can be encountered with an intermittent fault, for example, that appears randomly and irregularly for short time intervals. An incipient fault causing a slowly varying small-amplitude change can likewise have such impact on the process distribution. Obviously, a 'local' parameter reflecting a distribution property, i.e. mean, variance or kurtosis, is able to capture only a part of the change induced by these faults. It may fail to detect the fault presence unless the change is quite significant. Instead of monitoring 'local' parameters of probability distributions, it becomes more meaningful to monitor the overall shape of distributions in order to capture 'global' disparities and distortions. This makes call to informational measures defined between probability distributions. Comparing actual probability distributions against their healthy references using divergence measures allows to reveal the disparities caused by the faults onset.

The application of Kullback-Leibler (KL) divergence, as an informational measure, to detect and estimate small faults in multivariate processes, is investigated in this thesis. The sensitivity of KL divergence with respect to small changes is compared to other commonly used statistics for fault detection, namely $\mathbf{T^2}$ and SPE. The fault detection performance of the divergence is evaluated under the constraint of noisy environments, while referring to the Fault-to-Noise Ratio (FNR) as comparative criterion between the fault and noise levels.

As a first application, a 'global' time-domain analysis based on KL divergence and PCA is applied to eddy-currents testing (ECT) signals in order to reveal the signature of minor cracks and characterise their severity. The experimental data are acquired using an impedance sensor that scans several areas of a conductive plate having several thin cracks. Compared to the local statistical parameter, namely the mean statistic, the divergence is going to show a higher sensitivity to the imperceptible changes caused by the cracks presence.

A second application has been studied in this work to highlight the important role of the 'global' approach in FDD. Electrical machine faults with frequency signatures are studied, and validation is carried out on bearing faults. A global statistical analysis of the spectra of machine variables will be proved to lead to efficient features that are able to differentiate between different bearing faults according to the fault location and severity. Thanks to its 'global' character, the proposed method overcomes the *a priori* knowledge requirement of the characteristic fault frequencies. A particular attention is given to the ball-type faults since the corresponding frequency signature is very weak, which makes the discrimination among different severities of faults in bearing's balls challenging.

Contributions and outline

The principal contributions of this work are:

I- The application of KL divergence to the principal component variables in order to detect small faults in multivariate processes

- Comparison of the divergence sensitivity with respect to small changes with the sensitivity of commonly used fault indicators
- Evaluation, through detection error probabilities, of the detection performance in noisy environments

- Application of the divergence to the detection and characterisation of minor surface cracks in conductive plates, using experimental ECT signals

II- The estimation of the fault magnitude through an analytical model of the divergence

- Derivation, under some assumptions, of an analytical model of KL divergence that depends explicitly on the fault magnitude
- Model validation through simulation
- Evaluation of the estimation accuracy, for small faults and low fault-to-noise ratios (FNR), through the calculation of estimation errors
- Comparison with another estimation method

III- The exposition of a 'global' spectral analysis procedure for fault diagnosis in electrical rotating machines

- Extraction of efficient spectral features for discrimination among machine faults
- Validation on bearing faults using vibration signals
- Classification of bearing faults according to the fault location and severity

The present document is structured in 4 chapters.

In chapter 1, the literature related to the fault detection and diagnosis framework is reviewed and summarised. The main modelling approaches are briefly explained and discussed, showing their capabilities and limitations. A particular attention is paid to the data-driven approach due to its strong connection to the present work. The major data-driven methods are exposed in relatively more details.

Chapter 2, entitled KL divergence for fault detection, comprises the fault detection part of this work. The problem statement is described. The KL divergence is introduced along with its application to the principal components variables. Validation is carried out on simulated signals.

Chapter 3 is dedicated to the fault estimation. A theoretical study is conducted to obtain an analytical model of the divergence. An estimate of the fault amplitude is derived. Validation is carried out on a simulated AR process.

Chapter 4 comprises two parts. The first one concerns the application of KL divergence along with PCA to the detection and characterisation of minor surface cracks using experimental ECT data. The second one is dedicated to the application of a 'global' spectral analysis procedure for diagnosing faults in bearings of electrical machines.

Bibliography

[1] R. Isermann, "Supervision, fault-detection and fault diagnosis methods, an introduction," *Control Eng. Practice*, vol. 5, no. 5, pp. 639–652, 1997.

[2] P. M. Frank, "Analytical and qualitative model-based fault diagnosis, a survey and some new results," *European Journal of Control*, vol. 2, no. 1, pp. 6–28, 1996.

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

[4] T. Kourti, "Process analysis and abnormal situation detection: from theory to practice," *IEEE Control Systems*, vol. 22, no. 5, pp. 10–25, 2002.

[5] O. Bodnar and W. Schmid, "Cusum charts for monitoring the mean of a multivariate gaussian process," *Journal of Statistical Planning and Inference*, vol. 41, no. 6, pp. 2055–2070, 2011.

[6] C. Wikstrom, C. Albano, L. Eriksson, H. Friden, E. Johansson, A. Nordahl, M. S. Stefan Rannar, N. Kettaneh-Wold, and S. Wold, "Multivariate process and quality monitoring applied to an electrolysis process: Part i. process supervision with multivariate control charts," *Chemometrics and Intelligent Laboratory Systems*, vol. 42, no. 1-2, pp. 221–231, 1998.

[7] N. Martin, "Advanced signal processing and condition monitoring," *International Journal Insight on Non-Destructive Testing & Condition Monitoring*, vol. 49, no. 8, 2007.

[8] O. Ondel, G. Clerc, E. Boutleux, and E. Blanco, "Fault detection and diagnosis in a set "inverter-induction machine" through multidimensional membership function and pattern recognition," *IEEE Transactions on Energy Conversion*, vol. 24, no. 2, pp. 431–441, 2009.

[9] T. Boukra, A. Lebaroud, and G. Clerc, "Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 9, pp. 4034–4042, 2013.

[10] B. Trajin, J. Regnier, and J. Faucher, "Comparison between vibration and stator current analysis for the detection of bearing faults in asynchronous drives," *Electric Power Applications, IET*, vol. 4, no. 2, pp. 90–100, 2010.

[11] M. El Hachemi Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 984–993, 2000.

Chapter 1

Fault Detection and Diagnosis: State of the art

1.1 Introduction

The term 'fault' is commonly used to denote any unexpected behaviour (malfunction) of one of the system parts. A fault causes a small process plant deviation rather than a serious failure so it can be handled by fault tolerant control, of which fault diagnosis is a main function, to prevent undesired consequences. Typical examples of such faults are [1]:

- Actuator faults: faulty pumps, partial closing of valves, blocking of actuated joints or pipes, bearing faults, gear faults
- Sensor faults: biased measurements, drift, calibration errors, high percentage noise, dead zone
- Process (component, structural) faults: abnormal parameter changes, hardware defects like ruptures, cracks, loose parts

Faults, characterised with amplitude a, can also be classified according to the time-varying behaviour, as shown in Fig.1.1, into abrupt, intermittent and incipient (slowly varying) changes. According to the form the fault takes, it is sometimes useful to distinguish between discrete and distributed faults. In electromechanical systems, faults can either be mechanical or electrical.

The fault detection task requires that information about the normal (fault-free) behaviour of the plant/system is available. Information about the faulty behaviour

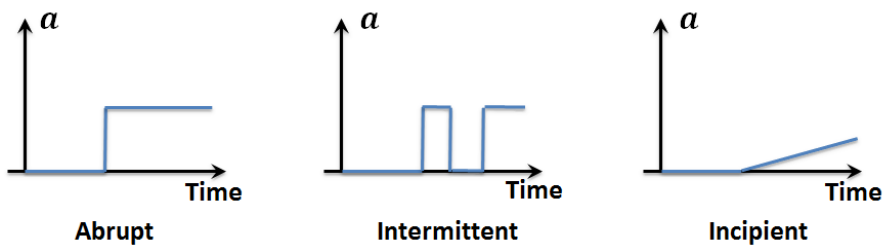

Figure 1.1: Fault types according to the time-varying behaviour

have also to be known if the fault diagnosis is of concern. This information can be of several forms:

- If the FDD scheme relies on an analytical model of the fault-free behaviour, then the faulty behaviour should likewise be modelled. Sensor faults for example affect the system outputs in an additive or multiplicative manner. Component faults are modelled as changes in the model parameters, and affect the process transfer function in a multiplicative manner if the system is linear.

- If the normal behaviour is described by a set of qualitative facts and rules compiled by the operator's expertise, then the information about the faults would likewise be in form of qualitative symptoms or qualitative states.

- Some faults can be recognised from characteristic features they introduce into measurements like spectral features, statistical moments, etc. One typical example is the case of faults (gear faults, bearing faults, short-circuit faults, broken rotor bars) in electromechanical systems.

Faults in plant equipment (actuators), instrumentation (sensors) and within the process itself are unavoidable due to age, misuse, harsh operating conditions and many other causes. They are responsible for the unwarranted change in the behaviour of the plant/system with respect to the nominal fault-free behaviour. Results could be at least reduction in performance and availability of processes, and at worst threats to the environment and human safety. It is extremely important to detect the early occurrence of a fault and to identify its severity and location to avoid the failing of the overall process by taking corrective control actions. In addition to actuator, sensor and process faults, other inputs to the plant can cause changes to the process behaviour without being harmful. Therefore, they could be

misleading to the fault detection task. These include noises, parameter variations, unknown disturbances and non-stationary conditions, commonly described as unknown inputs. The challenge for any FDD system is to achieve high sensitivity with respect to incipient faults while being robust against the unknown inputs.

The last three decades have testified the development of many advanced methods for early detection and diagnosis of incipient faults. The main principle behind is to compare system's actual behaviour against its nominal one to check for consistency. This can be achieved using a replication, 'redundancy', of the nominal behaviour of the system under monitoring. There are two major approaches to create redundancy [1]: either using a replication of hardware in order to compare outputs of identical components (hardware redundancy-based approach), or using a reconstruction/model of the system that estimates its nominal behaviour based on some *a priori* knowledge about the system (analytical redundancy-based approach). The hardware redundancy-based approach is extremely reliable but expensive and cumbersome and, therefore, its application is restricted to safety-critical processes as nuclear reactors, aeroplanes, etc. The analytical redundancy-based approach has been a major area of research since three decades or more. The methods developed in this framework can be classified into three main categories [2, 3]: quantitative model-based, qualitative model-based and process history based or so-called data-driven methods.

A comprehensive survey of fault detection and diagnosis methods is given in [4, 5]. Some review papers and books focus on analytical and qualitative model-based methods, see [6, 7, 8, 9] to cite a few of them. Data-driven methods are reviewed in [10, 11] among others. See [12, 13] for comparison of various methods.

This thesis revolves around signal processing employing data-driven methods for FDD. Three aspects of a fault diagnosis problem are addressed: the incipient fault detection, the fault magnitude estimation and the fault isolation. Attention is paid to the methodology which can subsequently be applied to a wide range of applications. Topics covered by the thesis are diverse in terms of tools and applications. This chapter briefly reviews the state of the art related to the context, describes the problems to be solved, states the contribution of this work and describes finally the outline of the remainder of the document.

1.2 Diagnosis process

The objective of fault diagnosis comprises three main tasks [4]:

- Fault detection, this is performed by designing fault indicators for which major goals are fast detection, sensitivity to incipient faults and robustness to various uncertainties.

- Fault isolation, it is the ability to distinguish between different faults. The problem is twofold. The simplest is to isolate faults occurring separately. The hardest is to be able to identify multiple faults occurring simultaneously.

- Fault identification, this is the determination of the time occurrence, the severity and the cause of a fault. This analysis is essential to predict the fault evolution (prognosis) and subsequently prescribes appropriate maintenance actions.

The realisation of these tasks depends decisively upon the information (amount, quality and nature) available about the system being monitored. Though the challenge for any diagnosis design is to achieve the desired objectives using minimum information and *a priori* knowledge. The available information is generally processed through the following diagnosis stages [6, 11]:

1. Residual/features generation/extraction. Residuals are signals designed to be sensitive exclusively to faults. They are ideally zero in fault-free conditions and minimally sensitive to various noises and disturbances. Features can be qualitative or quantitative in nature and they carry symptoms characterising the faults. For fault isolation, properly structured residuals or discriminative features are needed.

2. Residual/features evaluation/classification. The evaluation basically consists in a logical threshold testing. Furthermore, it can be viewed as a classification problem. This stage plays an important role in meeting the desired balance between performance (quick detection and isolation, low probability of missed detection) and robustness (low probability of false alarms, low classification error).

These two stages are root bases of fault diagnosis concept. The overall process is illustrated in Fig.1.2. A lot of researches have been carried out to propose different techniques to solve the residual/features generation/extraction and evaluation/classification problems. The techniques are guided by the prior knowledge or assumptions on the system being monitored. Most review and survey works classify the approaches of residual/features generation/extraction into three main categories, namely qualitative model-based, quantitative model-based and process

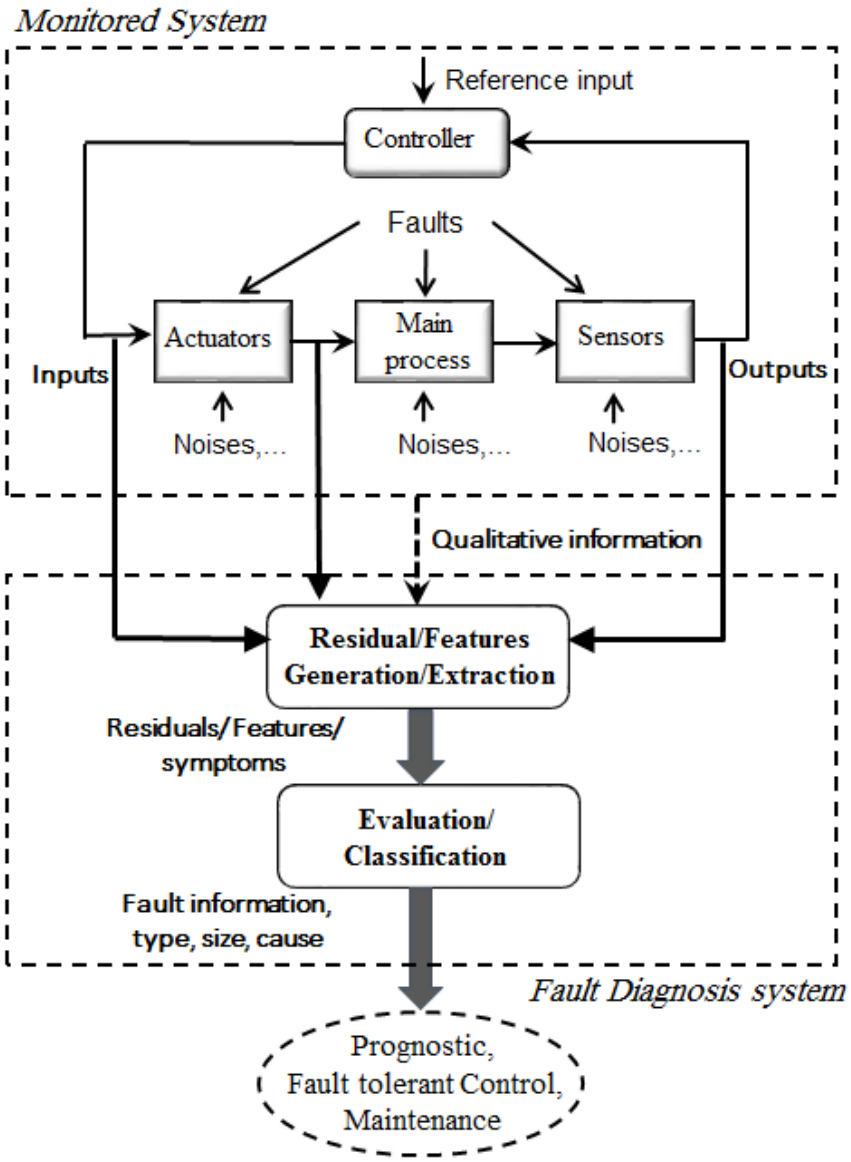

Figure 1.2: General Diagnosis Process

history-based. The quantitative and qualitative model-based methods use the concept of residual generation to perform fault detection and isolation. Process history based methods rely on feature extraction procedure. Statistical tools such as statistical classifiers, as well as non-statistical classifiers are proposed in the evaluation stage.

In the sequel, we briefly describe the basic principle of the model-based fault diagnosis using quantitative and qualitative models. We will then focus on the process history-based approach, which we have adopted for our work.

1.3 Quantitative and qualitative model-based approaches

From a model-based approach viewpoint, a fault can be detected if a reference or model of the nominal behaviour of the system being monitored is available. A model is commonly a knowledge representation. The knowledge can be quantitative (numerical) or qualitative (symbolic) in nature. Accordingly, the model can be mathematical, empirical, graphical, symbolic, explicit or implicit, etc. The variables observed from the system, consisting of sensor outputs, actuator inputs and process data, are basic representations of the system state. Modelling the fault-free system behaviour comprises the extraction of relationships among the observed variables, which are valid in normal operating conditions and change following a fault occurrence. The quantitative model-based approach expresses the relationships describing the system behaviour in terms of mathematical functions. The qualitative model in contrast expresses these relationships in terms of qualitative functions and heuristic symptoms. The quantitative approach uses analytical information to build the model, while the qualitative approach exploits as much knowledge about the process as possible and can easily handle qualitative information [6].

1.3.1 The quantitative approach

1.3.1.1 Quantitative model types

The behaviour of processes is governed by laws of physics. The quantitative approach aims at describing either the system's governing laws or the overall process behaviour (dynamics) using mathematical equations and functions over the observed variables. The challenge is to build a model as accurate, robust and complete as possible with a minimum modelling effort. According to the amount of information available to build such a mathematical model, it is possible to distinguish between three classes of mathematical models [5]:

- **White-box models:** They correspond to physical models based on a deep and concise understanding of the process physics. Sufficient information about the relationships among process variables and the functions describing their variations is available, yielding a comprehensive model for the system being

monitored. First principles models use established laws of physics to write equations relating the process variables without worrying about fitting the model to experimental data. A state-space model with known dynamics including real physical parameters is an example.

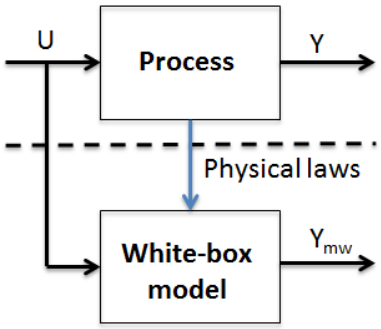

Figure 1.3: White-box model

Advantages: All the dynamics are known accurately.

Shortcomings: They may yield to complex models as systems are mainly multi-physics with coupled and non-linear phenomena.

- **Black-box (experimental) models:** This is the case of systems with unknown dynamics, providing hence little prior knowledge. A general input-output model aiming to only reproduce the system's output response to changes in its inputs can be used. Special structures for a linear input-output model are FIR model, ARX model, ARMAX model. Neural Network models are nonlinear black-box models which in real situations may achieve more accuracy than linear models. Black-box models have flexible structures whose parameters do not necessarily reflect physical meanings. The model is identified via system identification methods. A least square algorithm for example can be used to identify an ARX model, using observed data and the potential knowledge of the approximate orders and pure time-delay.

  Advantages: The identification step can be tuned to have simple models that fit in the objectives (control, diagnosis, optimisation, etc.)

  Shortcomings: They require the availability of experimental bench test, and the capability of injecting different excitation signals to cover all the system dynamics. Usually, the black-box models will have a limited domain of validity.

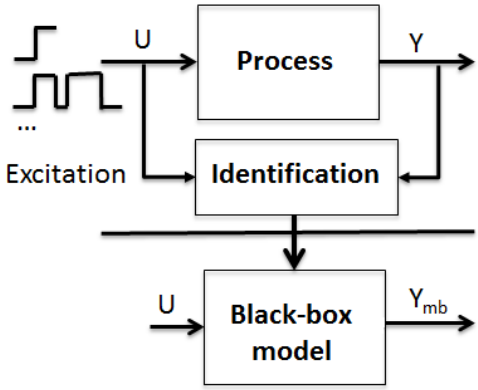

Figure 1.4: Black-box model

- **Grey-box (semi-physical or hybrid) models:** Such models are able to combine advantages from white-box and black-box models and overcome some of the problems of using either of them solely. This is the case when some physical grounds are available but not enough to construct a white-box model. The structure of the model can be partially identified from physical insights and some parameters should however be estimated from the observed data. The combination of a mathematical model representing the behaviour of the process and a neural network that compensates the model error is a grey-box model.

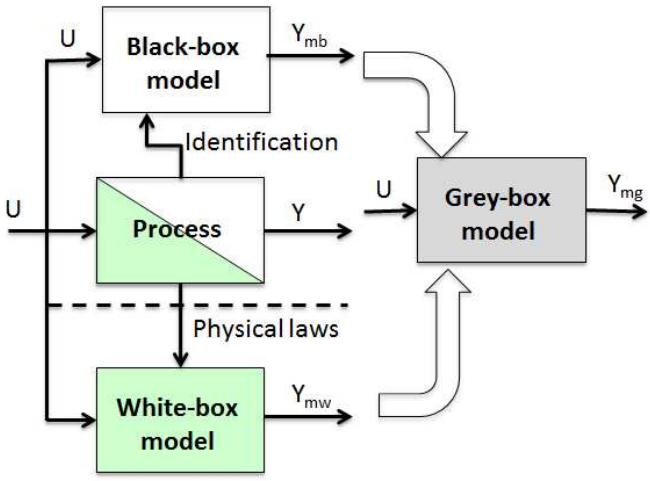

Figure 1.5: Grey-box model

⇒ They combine the advantages but also the drawbacks of white-box and black-box models.

1.3.1.2 Residual generation

Relying on the obtained fault-free quantitative model of the monitored system, the analytical methods that can be used for residual generation can be classified into three main categories:

- **Parity space approach:** Proposed in [14], it is especially suitable for actuator and sensor faults having an abrupt additive effect on the residual signal or vector. The basic idea consists in rearranging the input-output or state-space system equations to yield parity equations allowing for decoupling the residual from the system states and decoupling among different faults to make fault isolation. The robustness of residuals with respect to noises and unmodelled uncertainties is achieved by restricting the parity equations to those which are independent or at least weakly dependent upon them. This approach however can hardly cope with significant uncertainties, particularly the parametric multiplicative ones and the unknown disturbances. More details and improvements of the basic methods can be found in [15, 16, 17]. Examples of recent applications can be obtained in [18, 19].

- **Analytical observer-based approach:** It is based on application of output estimators designed as filters or observers and the use of resulting estimation error or a function of it as the residual [20, 21, 22, 23]. In fault-free conditions, the observer containing the model and an appropriate feedback will track the process so that the residual will only depend on the unknown inputs. When a fault occurs, the observer will model the system with less accuracy, thus increasing the magnitude of the residual. The main drawback is that incipient faults are difficult to detect. The fault isolation consists in using a bank of observers or filters each designed to be sensitive to a specific fault and robust with respect to the other faults and the unknown inputs. Such observers produce structured (enhanced) or directed residuals having total or approximate decoupling properties [24, 25, 26].

- **Parameter estimation approach:** It is mainly suitable for structural faults modelled as abrupt or slowly developing multiplicative changes in the model parameters. Residuals are generated by comparing the parametric model estimated on-line from acquired measurements against the reference model initially obtained in fault-free conditions [27, 28, 29, 30]. Changes in the model parameters which are potentially related to certain physical parameters are detected, providing a deep fault analysis. Parameter estimation methods such

as least squares and instrumental variables methods are used on-line; the approach becomes relatively intensive for large processes. Another disadvantage is that an input excitation is required in order to estimate the process parameters, causing problems for steady state operating plants.

These approaches are designed on the basis of an accurate and precise mathematical model of the system in its nominal operating mode. Therefore, the difficulty arises when dealing with complex, high-dimensional and nonlinear processes: on one hand developing a comprehensive model, if it is feasible, is costly, and on the other hand the more complex the system model is, the more tricky the robustness issue becomes. The application of the fault diagnosis approaches to a general non-linear system requires a model linearisation around the operating point, which can prove to be poor especially with systems that have severe non-linearities. Non-linear observers are designed for a restricted class of nonlinear systems [31, 32, 33, 34]. They work well for slowly varying faults, but fail to identify abrupt changes. Most of the techniques assume system linearity. Some works use black-box identification of some specific time-varying and/or nonlinear systems, with the aid of artificial neural networks (ANN) and adaptive ARX models for example [35, 36]. The development and validation of the black-box models need sufficient good experimental data, and thus abundance of experiments to be carried out. This might be impractical in real industrial processes.

As it has been mentioned before, the issue is to choose an accurate model for the system which is tractable and relevant for diagnosis. A lot of research has been carried out concerning the issue of complexity vs performance in quantitative model-based approaches. Important works are reviewed in [37, 38]. In large industrial systems for example, sometimes the best model would not be the detailed physical model (a white-box model) even if all system dynamics are known because this will yield a high-dimensional model which is heavy and complex from the diagnosis point of view. A small black-box parametric model identified from available data can however be useful for detection purposes. The problem here is whether it is still possible to infer diagnosis insights from the black-box model; parameters of a black-box model may lack of physical meanings. The robustness with respect to significant modelling uncertainties occurring due to parameter drifts poses a challenging problem for all the analytical model-based approaches [39, 40, 41].

1.3.1.3 Residual evaluation

The overall FDD quantitative model-based approach is illustrated in Fig.1.6. With the residual generation process, the complex problem of detecting and diagnosing

faults on a system turns into the problem of monitoring statistical properties of generated residuals. The goal is to decide whether the change as reflected by the residuals is significant indicating a fault occurrence or not. Afterwards, it would be interesting to estimate the change time and magnitude. This can be achieved using threshold logic [42], statistical decision theory [28, 43], pattern recognition [44, 45], fuzzy-decision making [46], or neural networks [47]. Threshold testing is the basic method and all the others can come down to a threshold test. A threshold should be suitably chosen so the false alarm's rate is minimised and the sensitivity to small faults is maximised. Time-variant adaptive thresholds have proven to be effective for a robust residual evaluation [48, 49].

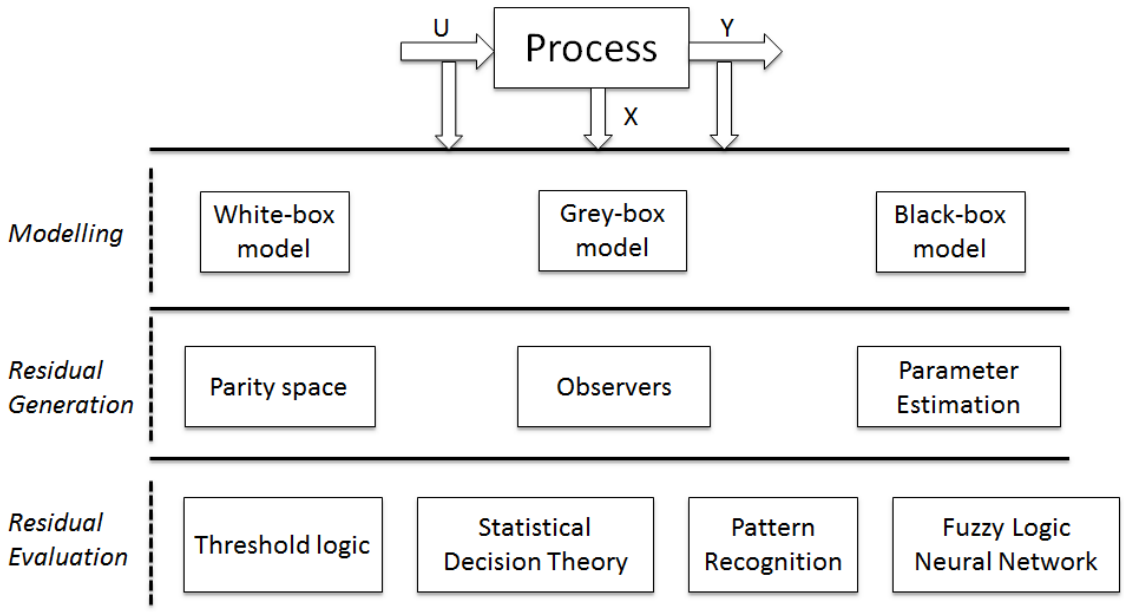

Figure 1.6: FDD quantitative model-based approach

To monitor the changes in the mean value or the spectral properties of the residuals, sophisticated statistical tools, namely sufficient statistics, have been derived from the likelihood ratio approach and/or the statistical local approach see, e.g. [28, 50, 51, 52]. Theoretical and practical studies have investigated the application of Cumulative sum (CUSUM) type algorithms, generalized likelihood ratio (GLR) test and modified version of them to change detection and to change time and magnitude estimation [53, 54]. When several residuals are generated so that each one reflects the occurrence of certain faults, using a bank of observers for example, decision can be made by feeding the set of residuals to a pattern recognition classifier.

1.3.2 The qualitative approach

1.3.2.1 Qualitative modelling

When the quantitative information available about the process dynamics in its nominal operating conditions is insufficient to build a high- fidelity mathematical model, the analytical fault diagnosis approaches described above would no longer be effective. Robust solutions can overcome this deficiency only to a certain extent and with great effort. The qualitative model-based approach, however, allows for exploiting as much knowledge about the process as possible and can easily handle qualitative information. A physical model of Fuel Cell systems for example is very complex because of coupling multiple physical domains (electrochemical, electrical, thermodynamics,...) [55]. Instead, a behavioral model describing the relationship among the qualitative variables is a good alternative/complement because it depends less on precise and detailed quantitative information [56, 57, 58]. In chemical industries [59, 60], the effect of a fault can be more easily described with qualitative means (heuristic symptoms represented as clauses or intervals) than quantitative ones such as a leakage of a pipe,... Besides, the qualitative modelling treats homogeneously linear and non-linear systems. The mathematically unmodelled dynamics (high order dynamics, high frequency oscillations, etc.) and non-linearities exhibited by some systems can be modelled using a behavioural description, see e.g. [56, 61, 62, 63].

Basically, qualitative information can be generated from process observations and previous experiences with abnormalities (faults) in a system. This leads to capturing the empirical relations between symptoms of faults and the faults themselves. A tree of if-then-else rules can be used to compactly represent (model) such relations. The acquired knowledge, referred to as 'shallow knowledge', is unfortunately uncertain and lacks process generality. It may fail if a novel condition not included in the knowledge base is encountered. It can however be useful for diagnosing small scale processes with limited knowledge. Otherwise, the diagnosis can be performed reliably by deriving 'deep knowledge' about the system behaviour using qualitative physics [6, 64, 65]. Accordingly, a set of qualitative equations termed as confluences, can be obtained from abstraction of the numerical values and the quantitative equations (differential equations), though the latter can be incomplete and imprecise. Confluence equations are able to describe qualitatively both the steady state and dynamic behaviour of a system. Causality graphical models, such as signed digraph (SDG), cause-effect graph (CE), constraint models developed from qualitative simulation (QSIM) algorithm are used to represent the cause-effect relationships expressed by the confluences [1, 66]. Compared to the quantitative model, the qualitative model is in general simpler, since some parameters in the quantitative model may not appear in the qualitative one. The qualitative model will therefore

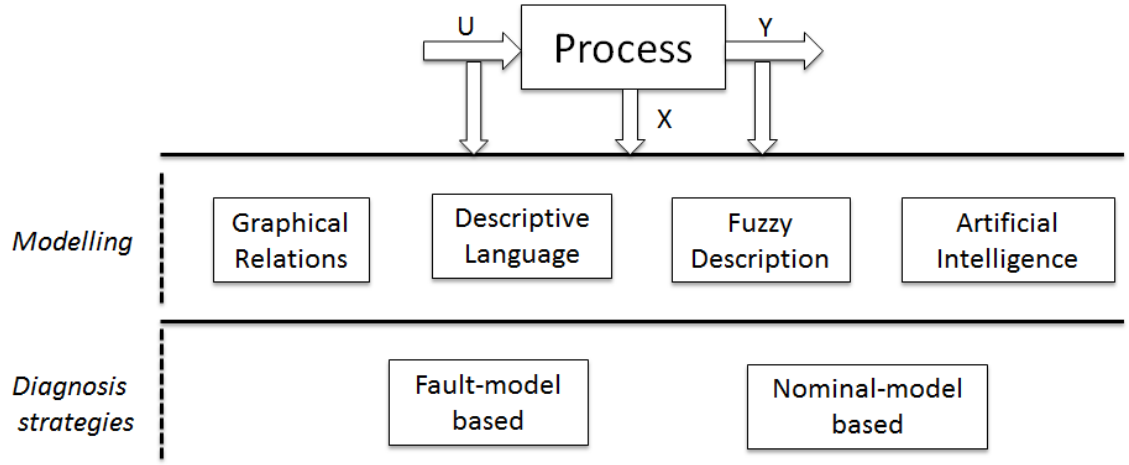

Figure 1.7: FDD qualitative model-based approach

be more robust to the uncertainties in these parameters.

The qualitative modelling starts with a qualitative description of signals in terms of global characteristics. Accordingly, the value of a signal at a given time is likely to satisfy a given condition such as to belong to a specific interval, or to be 'too high' or 'too low', etc. A set of similar conditions, referred to as symptoms, is defined for the system inputs, state variables and system outputs. This conversion of the numerical values to qualitative ones is responsible for the ambiguity problem inherent to the purely qualitative modelling approaches. Wrong diagnosis and false alarms could result from an unappropriated ambiguous symptoms setup. There are however several approaches able to reduce the ambiguity in purely qualitative reasoning, like the orders of magnitude reasoning approach [67], the semi-quantitative approach [68], the fuzzy quantity space approach[69], etc. The modelling thereafter proceeds by determining all possible changes in symptoms or qualitative values permitted to each variable and system parameter. The potentially incomplete quantitative model can be used to get the cause-effect relations among symptoms. Logic operators (AND, OR, XOR) like in fault trees, directed arcs like in SDG and algebra like in constraint models are used to represent the revealed changes and relations. The qualitative model normally grows rapidly with the behavioral complexity of the system, though there is no formal method to verify the accuracy of the model developed.

1.3.2.2 Diagnosis strategies

The different strategies for diagnosing the faults in a system with a qualitative model can be fundamentally classified into two broad categories [70]:

- Fault model-based, also referred to as symptomatic search strategy. It assumes that all fault modes and related symptoms are known *a priori* and modelled. The actual behaviour of the system is compared to the fault models. The predicted faulty behaviour matching the observed behaviour determines the fault or the set of faults that are present in the system. Such approach is advantageous for the fault identification and analysis, provided that all possible faults, even multiple faults and disturbances, are covered by the library of fault models.

- Nominal model-based, also referred to as topographic search strategy. Only nominal operating conditions of the system to be diagnosed are modelled. Any inconsistency between the observed behaviour and the reference behaviour reveals the presence of faults. No assumption on fault models are made, which means unpredictable faults can be detected. This approach is particularly useful when the system is decomposed into several subsystems for which the nominal behaviour is modelled in order to check their functionality.

Fig.1.7 displays the general FDD qualitative model-based approach. For details, the interested reader can refer to [64].

1.4 Data-driven approach

1.4.1 Background

In recent years, there has been an increasing interest in signal processing employing data-driven approaches in the context of condition monitoring, fault detection and diagnosis. The methods/tools proposed concern the analysis of process data in order to extract information for system health assessment and fault detection. They do not assume any form of model for the process under monitoring. They are especially useful when it comes to complex, large-scale and nonlinear processes, for which an accurate explicit model is unfeasible or expensive to develop or a tedious job, or the established model-based techniques are unsuitable. We also resort to these methods when the *a priori* knowledge about the process fundamentals and its behaviour is minimal and the only valuable resource is formed with historical process data. In

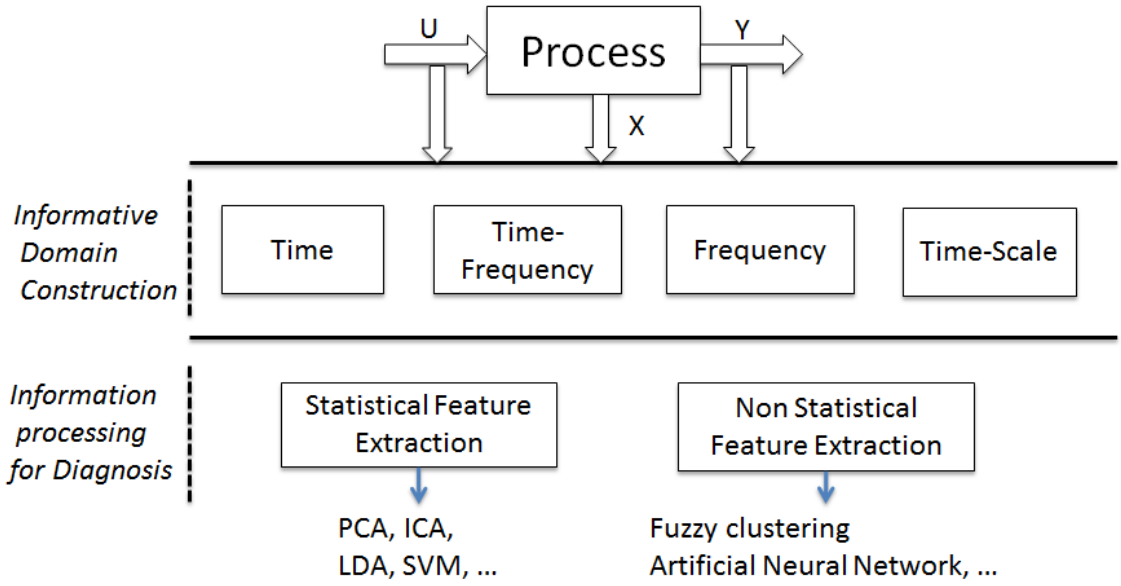

Figure 1.8: FDD data-driven approach

such cases, signals which can be of different types (mechanical, electrical, thermal measures and so on, input and output signals) are analysed in one or more informative domains (time, frequency, time-frequency and/or time-scale) to extract features sensitive to the fault presence, the fault severity and type.

The feature extraction process is mainly based on computational intelligence and machine-learning methods. These can be broadly classified into statistical and non-statistical feature extraction methods, see Fig.1.8. Multivariate statistical projection methods, including Principal Component Analysis (PCA), Partial Least Square (PLS), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), and statistical pattern classifiers are major components of statistical feature extraction methods. Neural Networks, K-means and fuzzy clustering approaches are non-statistical feature extraction methods. These are model-free powerful tools that can be used without being exhaustive for fault detection, fault visualisation, fault classification, identification, denoising, and therefore any FDD scheme makes use of these methods in a certain way. Application fields of these methods cover chemical, electrical, mechanical, aerospace, nuclear engineering processes.

The multivariate statistical projection methods are the most welcomed in industries compared to their data-driven alternatives, thanks to the satisfying performance to complexity ratio they achieve. A significant amount of researches have

been carried out on these methods in order to extend their capabilities to non-linear and dynamic processes, through many variants [71]. Recently, they have received a remarkable attention in the framework of electrical machine fault diagnosis. The detection and diagnosis results produced were similar or of better quality than those obtained with more complex methods, like neural networks or Support Vector Machines (SVM) for example. The main idea of these methods is to build low dimensional implicit models or spaces that capture intrinsic information from process data. This is achieved by finding linear combinations of variables (projections) that satisfy certain statistical properties. PCA, PLS and ICA extract geometric information which is generally useful for detection purposes. Other methods, namely statistical discriminant analysis [72], extract discriminative information useful for diagnosis and classification purposes. Whatever, the FDD scheme comprises two phases:

1. offline training: The main general tasks are construction of subspace model (feature extraction) from the process history data, threshold setup for detection, choice of discriminant function (boundaries) for classification, validation on test data.

2. on-line monitoring: The main general tasks are projection of the new measurements into the subspace model, calculation of test statistics (commonly $\mathbf{T^2}$ and SPE statistics) for detection and of discriminant function for diagnosis, comparison with threshold for making decision.

The main multivariate statistical projection methods are reviewed in the following, and an overview of their recent applications is given.

1.4.2 Subspace projection methods for feature extraction

1.4.2.1 Principal Component Analysis

Principal Component Analysis (PCA) refers to the problem of fitting a low dimensional linear subspace to a given set of measurements contained into a higher dimensional space. The data consist of measurements at N different times of some physical variables, for example temperature, voltage, vibration, etc, or artificial calculated variables (features) for example statistical moments. The time points play the role of observations. To solve this problem PCA uses linear correlations among variables. Statistically formulating the PCA problem, it consists in estimating the first principal components of a vector of random variables. The first principal components are defined to be uncorrelated linear combinations of the original variables that successively maximise the total variance of data projection [73]. From a geometrical viewpoint, this is equivalent to minimising the sum of squared error (euclidian

distance) between the observations and their projections. Both the geometrical and statistical formulation lead to the eigenvalue decomposition (EVD) as a solution, such that the first principal components are given by the dominant eigenvectors of the data covariance matrix, i.e. eigenvectors associated with the highest eigenvalues. PCA performs hence an orthogonal partition of the original variable/feature space into two complementary lower dimensional subspaces: the principal subspace, also called the signal or the representation subspace, which contains the most significant variations present in the data set and the residual subspace, also called the null subspace, which represents noises, redundancy, outliers (atypical observations, small-probability instances) and significant errors.

Such as it is defined above, PCA operates well for data which are at least approximately multivariate normal. It is for a multivariate normal distribution that the empirical sample covariance matrix is asymptotically an unbiased estimate of the true covariance matrix. Therefore if the normality condition is satisfied, the principal components, which are estimated as the eigenvectors of the sample covariance matrix, will be also unbiased estimate of the true principal components. Under this assumption, statistical inferences about the underlying population of principal components can be derived analytically [74]. However the numerous applications of PCA, especially in data compression and dimensionality reduction, show that the multivariate normality of data is not a strict assumption and PCA still can be applied as descriptive tool even if the statistical nature of the data is unclear. Some works have even gone toward reducing the effect of atypical data observations which do not belong to the main structure of data but can influence the result of PCA. Robust estimation methods of principal components either directly from data or via a robust estimation of the covariance matrix are proposed in order to tolerate the presence of such extreme observations. These methods require a careful examination of any observations that have been omitted or significantly down-weighted by the analysis, and they can be time consuming. See [75, 76] for a detailed discussion on these attempts.

PCA can be applied in many situations where the form of probability distribution of data slightly deviates from normal. It may unfortunately fail when the second-order statistics (the mean and the variance) do not reflect the data information because the normality assumption is severely violated. In Probabilistic PCA (PPCA), the noise is assumed to be drawn from an unknown distribution, and it is found that a non-gaussian noise leads to a principal subspace which is no longer affine. Another extension of PCA is the kernel PCA (KPCA), which tackles the problem of identifying a nonlinear manifold from sample data [77, 78]. The main difficulty with this approach is the choice of a good kernel. The kernel should be

chosen according to the structure of the nonlinear manifold to be identified, though the exact structure is not available. The state of the art however reviews numerous applications of KPCA to the fault detection and diagnosis of various systems and to process monitoring; this primarily owes to the fact that KPCA can handle a wide range of non-linearities and it is simpler than its alternatives based on neural networks or genetic programming [79].

Another limitation of classical PCA is the assumption of minimal auto-correlation in each variable (data are independently distributed). This assumption can be violated in chemical processes for example where the measurements are correlated over time (time series data), and in many other cases due to process dynamics and high sampling frequency as instance. The PCA in its classical form does not account for the non-zero auto-correlation at high time lag, and this generally causes too many false alarms and wrong decisions in the used PCA-based detection control charts. Therefore, variants to PCA like dynamic PCA (DPCA) have been proposed to tackle the dynamic issues [80, 81]. The idea of DPCA is to perform PCA on a time-lagged data matrix, thus identifying both the linear static and dynamic relationships among variables. On the other hand, the computed principal components are time-invariant and the process generating data is considered to run under stationary operating conditions. If the process is non-stationary, the data statistics will be time-varying and PCA should be updated to learn the time-varying principal components [82, 83].

1.4.2.2 Partial Least Squares

Partial Least Squares (PLS), or projection to latent structures, and PCA are the basic multivariate projection methods used in multivariate statistic process monitoring. PCA analyses the overall correlation structure of the process variables and monitors all the variations. PLS divides the process variables into observed (input or measurement) and unobserved (output or quality) ones and attempts to extract from the observed variables latent variables which are the most correlated and relevant to the unobserved variables. Therefore, by analysing the inputs-outputs correlation, PLS allows monitoring among all the variations in process variables only those which affect the output variables. There are several algorithms to calculate PLS-based latent variables, the most instructive method is known as NIPALS for Non-linear iterative partial least squares algorithm. We refer the reader to the chemometrics literature [84, 85] for details on PLS algorithms and applications.

Data-driven models from PLS have been mainly applied in multivariate statistical quality control to monitor output quality variables by monitoring input process

variables. Beside monitoring, they are used for quality prediction and optimisation. In this framework, several PLS variants have been proposed to overcome some shortcomings of the classical PLS, the most recent ones are total projection to latent structures (TPLS) [86, 87] and concurrent PLS (CPLS) [88, 89]. Kernel PLS (KPLS) was proposed, similarly to KPCA, to handle non-linear processes [90]. Applications are mainly in industrial manufacturing and production processes, including chemicals, polymers, and microelectronics. In situations where it is meaningless to consider the separation between variables, i.e. there is no quality variables and all variables are treated homogeneously, PCA should be used as being conceptually equivalent to PLS. However, we still can find some applications of PLS in conjunction with other techniques to make faults classification and diagnosis [91, 92, 93].

1.4.2.3 Independent Component Analysis

Independent Component Analysis (ICA) is, like PCA, a statistical method for capturing the essential information from data sets of random variables, measurements or features. PCA, by looking for uncorrelated linear combinations of the observed variables, accounts for only the second-order statistics of data. ICA however reveals information on higher-order statistics, and looks for linear combinations which are not only uncorrelated but also independent. Data from m variables are assumed to be a mixing of l independent latent (unobserved) components ICs (the source). The ICs are random variables with unknown non-gaussian distribution. Such case is encountered, for example, in mechanical systems where several sensors are placed in different positions so that each sensor measures a mixture of stochastic vibration signals emitted by different parts of the system. This phenomenon is described by the equation $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{x}$ is the random $(m \times 1)$-vector whose elements are the sensor measurements, $\mathbf{s}$ is the random $(l \times 1)$-vector whose elements are the original independent vibration components, generally unmeasurable, and $\mathbf{A}$ is the mixing $(m \times l)$-matrix which is also unknown. The literature provides various algorithms of ICA that attempt to estimate $\mathbf{A}$ and $\mathbf{s}$ based on the general assumption of statistical independence. The interested reader can refer to [94, 95] for more details.

ICA, by extracting the directions that are as statistically independent as possible and projecting the process data onto the associated basis vectors, can be used for dimensionality reduction, data compression, denoising and the extraction of non-redundant relevant features enhancing the fault diagnostics routine. It has been applied in dynamic chemical complex processes in order to capture non-gaussian features that proved to be effective for fault detection and diagnosis [96, 97, 98]. In addition, ICA finds numerous applications in electrical machine fault diagnosis and monitoring of power system. It has been employed to identify signatures of different faults in induction motors (broken rotor bars, bearing faults) using stator current

signals and vibration measurements [99, 100, 101]. The ICA-based signatures have been used therein to monitor the motor condition and perform online fault detection through the integration of a classification system, neural network or SVM. It has also been applied to gearbox condition monitoring, particularly to remove the noise from vibration signals and detect the signal transients [102, 103]. ICA has been used in [104] to separate the transient oscillations from the steady-state variations in the voltage and current signals in power system, allowing the classification of disturbances and monitoring the power quality. See also, [105]. In [106], independent components of currents have served to derive detection index of power system faults.

1.4.2.4 Linear Dicriminant Analysis

Linear Dicriminant Analysis (LDA), also known as FDA for Fisher Discriminant Analysis, is a popular multivariate statistical method for solving pattern recognition problems. However, it is only recently that LDA has been applied to electrical machine fault diagnosis [107, 108]. LDA shares a common character with PCA: they analyse and capture second-order statistics, and thus they produce optimal results for normally distributed data. Healthy (faultless) data are generally processed with PCA to identify region of normal operation and perform fault detection. LDA processes healthy and faulty data simultaneously and aims at extracting discriminative directions from overlapping data. It is therefore meaningful in fault diagnosis. The database for each fault class to be considered can be acquired from the real process during the corresponding faulty conditions, or generated from a simulation of the process in healthy and faulty operating modes. The fault diagnosis based on statistical discriminant analysis comprises two tasks :

1. Extraction of discriminative directions leading to maximum separation between known fault classes,

2. Choosing a discriminant function leading to minimum classification errors.

LDA defines optimal discriminative directions as linear combinations of the original variables which maximise the ratio between the inter-class dispersion and the intra-class dispersion. Lagrange multipliers lead to the eigenvalue decomposition (EVD) of a given symmetric matrix as a solution. The eigenvectors associated with the non-zero eigenvalues define discriminative axes which generate the LDA space that will be used for classification. For any c class problem we would always have $c-1$ non-zero eigenvalues. Subsequently, the dimension of the original space can be largely reduced using LDA, since the new dimension depends only upon the number of classes. Thus, the projection of the original data onto the LDA space allows the separability and also visualisation of classes . Afterwards, a discriminant function must be expressed in terms of the discriminant variables, i.e. the original variables

after projection, in order to classify and diagnose a new observation. LDA uses linear function, a line or a plane. Non-linear functions that help reducing miss-classifications can also be obtained using non-linear discriminant analysis.

Several variants to LDA have been investigated to overcome the drawbacks of the classical method. One major drawback is rather computational; the intra-class scatter should be nonsingular and some approaches have been presented to address this problem [109]. Another shortcoming is that samples in a class are supposed to be unimodal, i.e. they form one cluster. In situations where the data to be processed have complex distributions so that samples in one class are separated in several clusters, LDA gives poor results. Local Fisher Discriminant Analysis (LFDA) has been proposed to address the issue of intra-class multimodality [110]. Besides, the sensitivity of LDA to outliers and extreme samples has been addressed, and a linearly optimized discriminant analysis (LODA) was proposed to obtain robust LDA results [111]. Nonlinear extension of FDA, namely kernel FDA (KFDA), has been also proposed and applied to nonlinear classification [112].

Despite being old, originally developed by Fisher in 1936, it is only recently that LDA has been applied to electrical machine fault diagnosis. In [113] LDA has been used along with PCA to detect and classify broken rotor bars in induction machines. M. Sahni and al. present in [114] the application of LDA to classify the severity of the coil arcing fault in low-voltage motor. The authors in [115] use LDA classification to detect the type of fault in permanent magnet synchronous machine, particularly static eccentricity, inter-turn short circuit and demagnetization faults, and to identify the severity of eccentricity fault. Authors in [116] show the detection of gearbox faults. The trace ratio linear discriminant analysis (TR-LDA) has been lately proposed for classification of high dimensional non-gaussian fault data [117]. The method was applied for visualisation and classification of bearing fault data using features extracted from vibration signals. The faulty data include data from single point defects, surface roughness, and from inner race faults, outer race faults and ball faults. The results showed superiority of TR-LDA over other methods, especially PCA, canonical variate analysis (CVA) and classical LDA.

1.4.2.5 Other subspace methods

Other subspace methods that were applied to process monitoring are singular spectrum analysis (SSA) whose idea is similar to dynamic PCA, canonical variate analysis (CVA) [118, 119] which is similar to PLS, canonical discriminant analysis (CDA) [120] which is related to PCA and CVA. Many high-resolution parametric spectral estimation techniques that have been used in electrical machine FDD are based on the subspace decomposition between signal and noise. These include SSA, MUSIC

(multiple signal classification) and ESPRIT methods (estimation of signal parameters via rotational invariance technique). These methods are powerful at extracting the true spectral information from signals with low signal-to-noise ratios, thus improving the detection of harmonic components associated with motor faults. This is rendered possible by converting the fault detection problem into a generalised eigenvalue decomposition problem. Applications cover detection of rotor broken bar fault [121], identification of multiple faults [122] and tool wear monitoring [123]. Very recently, the MUSIC has been applied to grid diagnostics [124], the ESPRIT to broken rotor bar fault detection [125] and SSA to rolling element bearing fault diagnosis [126].

1.5 Synthesis

This chapter is an overview of the state-of-the-art for FDD approaches. It explains the principles of the different methods and shows their capabilities and limitations. The general problem of FDD is resolved through three main stages: the modelling

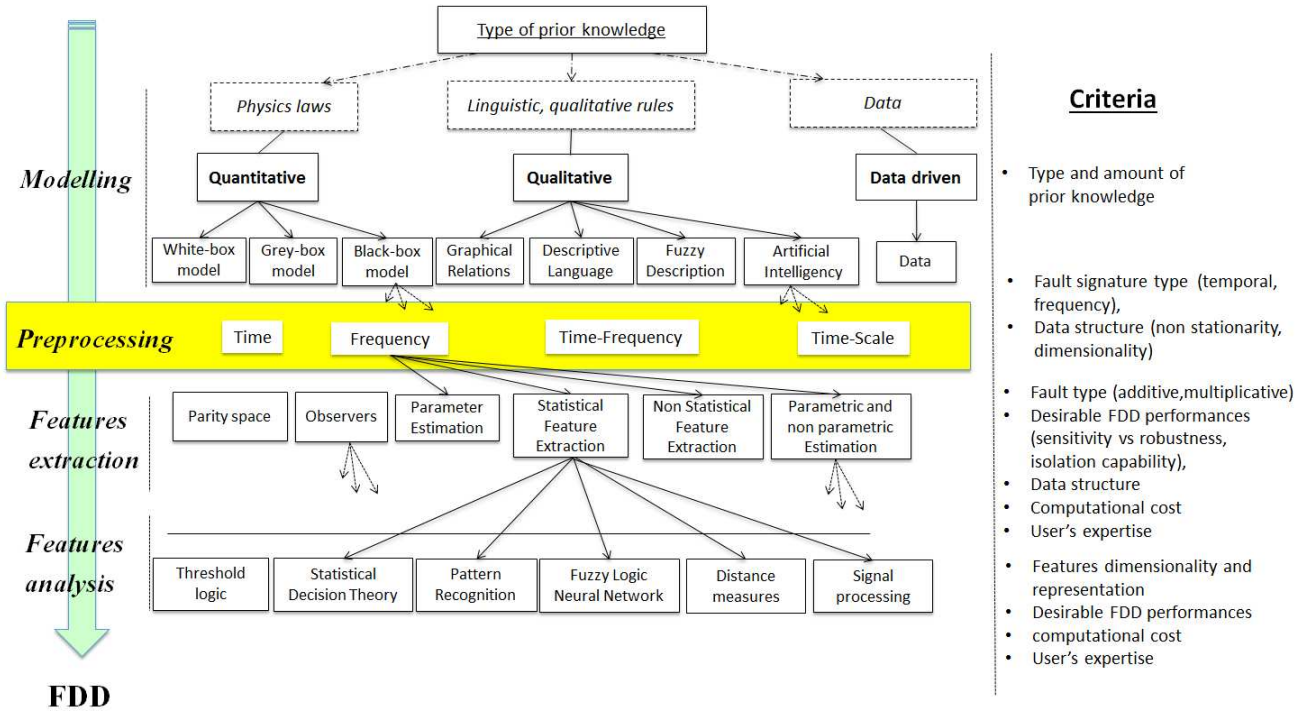

Figure 1.9: A three-stage FDD process

(problem formulation), the features extraction and the features evaluation. A preprocessing intermediary stage precedes the features extraction and can performs

tasks like features selection, signals filtering or shaping, signals transformation into informative domains (frequency,etc.), ... Fig.1.9 summarizes the different possible approaches in the 3-stage process. It points out the main criteria that differ the different approaches, and that should be considered by a neophyte, who is looking for the FDD procedure that is suitable for his application.

1.6 Problem statement

Except the intermittent faults, faults are classified into two categories: abrupt and incipient faults. Abrupt faults are manifest changes in the measured data like step or impulse faults. They normally lead to pronounced mean or variance shifts. To the best of our knowledge, there is no strict definition for incipient faults. However, an incipient fault is usually shown as a slowly varying drift change since it is not abrupt. A small intermittent change of short duration is also incipient. Incipient faults can be caused for example by material premature ageing due to harsh operating conditions or misuse. A sensor gain drift of some percent ($\simeq$ 1% to 10%) and a pitch of 0.18 mm on a ball bearing with a diameter of 8 mm (thus corresponding to a $0.18/8 = 2\%$ fault) are some practical examples. However, if not detected, such faults can lead in the short or medium-term to an abrupt fault inducing the system failure.

We may consequently consider incipient any imperceptible fault whose signature is likely masked by normal process variations and noises. Unfortunately, such fault might not result in a mean or variance shift. It however certainly induces distortions and some change in the probability distribution of observations of process variables. Probability distributions are commonly used to represent graphically the behaviour of parameters and variables in a process, through histograms and probability plots. Parameters of probability distributions, particularly the mean, the variance and sometimes the higher order moments can be computed to numerically monitor the process behaviour. Most statistics for faults detection are designed to detect shifts of mean or variance in the process variables. Nevertheless, it seems relevant to assess the overall shape of probability distributions for the detection of incipient faults, and without assuming a specific form or model for the distribution. Toward this end, Kullback-Leibler (KL) divergence is proposed as a fault indicator taken from informational theory. This measure has been previously applied for anomaly detection and pattern recognition problems in many fields. The present study investigates the detection and estimation of small faults in multivariate processes using the KL divergence measure. So the divergence will be compared to statistic tests usually used for fault detection with the subspace projection methods, namely the Hotelling $\mathbf{T^2}$ and SPE.

The detection of incipient faults is particularly challenging in the framework of structural health monitoring that uses non-destructive inspection techniques. Many engineering systems are subject to faults that start with minor material cracks, but can rapidly develop into a serious damage leading to a catastrophic failure. Ensuring the integrity of structures is critical for applications like aeroplanes, ships, submarine, nuclear reactors, etc. The Eddy-Current Testing (ECT) is the most commonly used nondestructive evaluation technique thanks to the benefits it provides, in terms of the fast and simple testing, the sensor portability and the suitability to complex geometrical structures. The difficulty to detect minor cracks with the ECT is due to the fact that eddy currents are responsive to the presence of cracks, but also to many other noise sources that may obscure the weak signature of cracks. These noise sources include the environmental perturbations like temperature variations, the lift-off variations, the surface roughness, etc. In this context, a global time-domain analysis based on KL divergence and PCA will be applied to ECT signals with low SNR in order to reveal the signature of minor cracks and characterise their severity.

Beside the fault detection and estimation problems, some problems related to fault diagnosis, particularly in electrical rotating machines, are also tackled in this study. Mechanical faults, such as rotor eccentricity, bearing faults, shaft misalignment and load faults are common failures in motors. They must be detected at their inception to avoid down time and even damage to other related machinery. They are known to have characteristic frequency signature. The magnitude of some harmonic components associated to the fault type increases with the fault presence. Therefore, the detection of a particular fault is based on monitoring the energy level at the associated characteristic frequencies or into some bands related to them. Usually the characteristic fault frequencies must be known, or estimated, in advance. Bearing faults in particular are the most frequently encountered in rotating machines. The incipient bearing faults are in form of minor dents, spalls or a crack which may appear in either one of the bearing elements, namely the inner race, the outer race, the balls or the cage. Identifying the spall's location is important because it gives insights about the fault causes. The characteristic frequency associated with each bearing element can be estimated through theoretical kinematic expressions using the specifications of the bearing and the rotational speed of the motor.

The difficulty of fault diagnosis in motors arises when the actual characteristic fault frequencies differ from the theoretical estimated ones due to various factors such as ageing, unexpected rotor slip, excessive friction, fluctuations of rotor speed, etc. In addition, in real situations and under some conditions of load distribution

and rotor speed, the fundamental characteristic frequencies may not be present into the signal spectrum although the fault is quite serious. They can be weak, or even masked by the noise and other stronger machine frequencies. Instead, usually several multiple harmonics of these frequencies and sidebands related to their modulations by the rotational frequencies are almost always pronounced. Therefore, although the ability to detect the characteristic fault frequencies is necessary, this approach neglects a large part of the total signature lying throughout the spectra of machine variables. The difficulty with the overall frequency signature is that the occurrence of a particular harmonic or a sideband is a random process and its level depends on the fault itself and likewise other unpredictable causes such as the operating load condition and the presence of other machine problems like load imbalance, shaft misalignment, looseness, etc. In consequence, we propose a 'global' spectral analysis procedure able to capture the entire spectral signature of faults. An important advantage of the procedure is that it overcomes the usual *a priori* knowledge requirement of the characteristic fault frequencies.

1.7 Conclusion

This chapter has covered the main FDD approaches. The data-driven fault diagnosis has an important advantage over the model-based one: it can handle the data noises and the non-stationary features easily with advanced signal processing and data analysis techniques. The ability to detect and diagnose faults at early stage of development is still an issue for all approaches. In the presence of high-level noise and non-stationary features like the process parameter variations for examples, this issue is more tricky because these unknown inputs can mask the fault signature, or also affect the decision making task. A PCA-based KL divergence approach for incipient fault detection will be discussed is the next chapter.

Bibliography

[1] R. J. Patton, P. M. Frank, and R. N. Clark, *Issues of Fault Diagnosis for Dynamic Systems*. Berlin Heidelberg: Springer, 2000.

[2] V. Venkatasubramanian, R. Raghunathan, Y. Kewen, and N. K. Surya, "A review of process fault detection and diagnosis part I: Quantitative model-based methods," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.

[3] D. Xuewu and G. Zhiwei, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2226–2238, 2013.

[4] R. Isermann, "Supervision, fault-detection and fault diagnosis methods, an introduction," *Control Eng. Practice*, vol. 5, no. 5, pp. 639–652, 1997.

[5] ——, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Berlin Heidelberg: Springer, 2006.

[6] P. M. Frank, "Analytical and qualitative model-based fault diagnosis, a survey and some new results," *European Journal of Control*, vol. 2, no. 1, pp. 6–28, 1996.

[7] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–53, 2010.

[8] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems*. Boston, Massachusetts: Kluwer Academic Publishers, 1999.

[9] S. Baniardalani, J. Askari, and J. Lunze, "Qualitative model based fault diagnosis using a threshold level," *International Journal of Control, Automation and Systems*, vol. 8, no. 3, pp. 683–694, 2010.

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

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

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

[13] P. P. Harihara, K. Kyusung, and A. G. Parlos, "Signal-based versus model-based fault diagnosis: A trade-off in complexity and performance," *4th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 277–282, 2003.

[14] M. Sturza, "Navigation system integrity monitoring using redundant measurements," *Journal of the Institute of Navigation*, vol. 35, no. 4, pp. 69–87, 1988.

[15] R. J. Patton and J. Chen, "Review of parity space approaches to fault diagnosis for aerospace systems," *Journal of Guidance Control and Dynamics*, vol. 17, no. 2, pp. 278–285, 1994.

[16] Y. Hao, W. Guizeng, and S. X. Ding, "A new parity space approach for fault detection based on stationary wavelet transform," *IEEE Transactions on Automatic Control*, vol. 49, no. 2, pp. 278–287, 2004.

[17] P. Beckerle, H. Schaede, N. Butzek, and S. Rinderknecht, "Balancing filters: An approach to improve model-based fault diagnosis based on parity equations," *Mechanical Systems and Signal Processing*, vol. 29, pp. 137–147, 2012.

[18] H. Berriri, M. W. Naouar, and I. Slama-Belkhodja, "Easy and fast sensor fault detection and isolation algorithm for electrical drives," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 490–499, 2012.

[19] S. Guerrier, A. Waegli, J. Skaloud, and M. P. Victoria-Feser, "Fault detection and isolation in multiple mems-imus configurations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2015–2031, 2012.

[20] J. Wunnenberg, "Observer-based fault detection in dynamic systems," Ph.D. dissertation, Dissertation Universitat Duisburg, Fachgebiet Me-β und Regelungstechnik, Fortschritts-berichte VDI, Reihe 8, Nr 222, VDI-Verlag Dusserldorf, 1990.

[21] T. A. Najafabadi, F. R. Salmasi, and P. Jabehdar-Maralani, "Detection and isolation of speed-, DC-link voltage-, and current-sensor faults based on an adaptive observer in induction-motor drives," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1662–1672, 2011.

[22] G. H. B. Foo, Z. Xinan, and D. M. Vilathgamuwa, "A sensor fault detection and isolation method in interior permanent-magnet synchronous motor drives based on an extended Kalman filter," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3485–3495, 2013.

[23] F. Pierri, G. Paviglianiti, F. Caccavale, and M. Mattei, "Observer-based sensor fault detection and isolation for chemical batch reactors," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 8, pp. 1204–1216, 2008.

[24] P. M. Frank and X. Ding, "Frequency domain approach to optimally robust residual generation and evaluation for model-based fault diagnosis," *Automatica*, vol. 30, no. 5, pp. 789–804, 1994.

[25] X. J. Li and G. H. Yang, "Dynamic observer-based robust control and fault detection for linear systems," *IET Control Theory & Applications*, vol. 6, no. 17, pp. 2657–2666, 2012.

[26] C. Wen, T. C. Wei, M. Saif, F. L. Meng, and W. Hai, "Simultaneous fault isolation and estimation of lithium-ion batteries via synthesized design of luenberger and learning observers," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 290–298, 2014.

[27] R. Isermann, "Fault diagnosis of machines via parameter estimation and knowledge processing-tutorial," *Automatica*, vol. 29, no. 4, pp. 815–835, 1993.

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

[29] V. Reppa and A. Tzes, "Fault detection and diagnosis based on parameter set estimation," *IET Control Theory & Applications*, vol. 5, no. 1, pp. 69–83, 2011.

[30] B. Bellali, A. Hazzab, I. K. Bousserhane, and D. Lefebvre, "Parameter estimation for fault diagnosis in nonlinear systems," *Procedia Engineering*, vol. 29, pp. 2016–2021, 2012.

[31] W. T. Chen and M. Saif, "Actuator fault diagnosis for a class of nonlinear systems and its application to a laboratory 3d crane," *Automatica*, vol. 47, no. 7, pp. 1435–1442, 2011.

[32] ——, "sliding mode observer-based strategy for fault detection, isolation, and estimation in a class of lipschitz nonlinear systems," *International Journal of Systems Science*, vol. 38, no. 12, pp. 943–955, 2007.

[33] T. F. Lootsma, "Observer-based fault detection and isolation for nonlinear systems," Ph.D. dissertation, Institut for Elektroniske Systemer, Aalborg Universitet, 2001.

[34] J. Zhang, A. K. Swain, and S. K. Nguang, "Detection and isolation of incipient sensor faults for a class of uncertain non-linear systems," *IET Control Theory & Applications*, vol. 6, no. 12, pp. 1870–1880, 2012.

[35] Q. Zhang, "Using nonlinear black-box models in fault detection," *Proceedings of the 35th IEEE Conference on Decision and Control*, vol. 1, pp. 636–637, 1996.

[36] L. B. Palma, F. V. Coito, and R. N. Silva, "Fault diagnosis based on black-box models with application to a liquid-level system," *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation*, vol. 2, pp. 739–746, 2003.

[37] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601–611, 1976.

[38] C. T. Chang and J. I. Hwang, "Simplification techniques for EKF computations in fault diagnosis-suboptimal gains," *Chemical Engineering Science*, vol. 53, no. 22, pp. 3853–3862, 1998.

[39] R. J. Patton and J. Chen, "Robust fault detection and isolation (FDI) systems," *Control and Dynamic Systems*, vol. 74, pp. 171–224, 1996.

[40] X. Wu and G. Campion, "Fault detection and isolation of systems with slowly varying parameters-simulation with a simplified aircraft turbo engine model," *Mechanical Systems and Signal Processing*, vol. 18, no. 2, pp. 353–366, 2004.

[41] S. Tornil-Sin, C. Ocampo-Martinez, V. Puig, and T. Escobet, "Robust fault diagnosis of nonlinear systems using interval constraint satisfaction and analytical redundancy relations," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 1, pp. 18–29, 2014.

[42] A. Emami-Naeini, M. M. Akhter, and S. M. Rock, "Effect of model uncertainty on failure detection: the threshold selector," *IEEE Transactions on Automatic Control*, vol. 33, no. 12, pp. 1106–1115, 1988.

[43] C. Svard, M. Nyberg, E. Frisk, and M. Krysander, "Data-driven and adaptive statistical residual evaluation for fault detection with an automotive application," *Mechanical Systems and Signal Processing*, vol. 45, no. 3, pp. 170–192, 2014.

[44] A. W. Osburn, T. M. Kostek, and M. A. Franchek, "Residual generation and statistical pattern recognition for engine misfire diagnostics," *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 2232–2258, 2006.

[45] R. Razavi-Far, E. Zio, and V. Palade, "Efficient residuals pre-processing for diagnosing multi-class faults in a doubly fed induction generator, under missing data scenarios," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6386–6399, 2014.

[46] H. Sneider and P. M. Frank, "Observer-based supervision and fault detection in robots using nonlinear and fuzzy logic residual evaluation," *IEEE Transactions on Control System Technology*, vol. 4, no. 3, pp. 274–282, 1996.

[47] B. Koppen-Seliger, P. M. Frank, and A. Wolff, "Residual evaluation for fault detection and isolation with RCE neural networks," *Proceedings of the American control conference*, vol. 5, no. 3, pp. 3264–3268, 1996.

[48] P. M. Frank, "Residual evaluation for fault diagnosis based on adaptive fuzzy thresholds," *IEE Colloquium on Qualitative and Quantitative Modelling Methods for Fault Diagnosis*, pp. 401–411, 1995.

[49] Z. Xiaodong, M. M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 4, pp. 576–593, 2002.

[50] M. Basseville, "On-board component fault detection and isolation using the statistical local approach," *Automatica*, vol. 34, no. 11, pp. 1391–1415, 1998.

[51] G. Rigatos and Q. Zhang, "Fuzzy model validation using the local statistical approach," *Fuzzy Sets and System*, vol. 160, no. 1, pp. 882–904, 2009.

[52] F. Kiasi, J. Prakash, S. Patwardhan, and S. L. Shah, "A unified framework for fault detection and isolation of sensor and actuator biases in linear time invariant systems using marginalized likelihood ratio test with uniform priors," *Journal of Process Control*, vol. 23, no. 9, pp. 1350–1361, 2013.

[53] F. Gustafsson, "Statistical signal processing approaches to fault detection," *Annual Reviews in Control*, vol. 31, no. 1, pp. 41–54, 2007.

[54] I. Nikiforov, V. Varavva, and V. Kireichikov, "Application of statistical fault detection algorithms to navigation systems monitoring," *Automatica*, vol. 29, no. 5, pp. 1275–1290, 1993.

[55] F. Barbir, "Chapter eight - fuel cell diagnostics," in *PEM Fuel Cells, Theory and Practice (Second Edition)*, 2013, pp. 265–304.

[56] Z. Zheng, R. Petrone, M. Pera, D. Hissel, M. Becherif, C. Pianese, N. Yousfi Steiner, and M. Sorrentino, "A review on non-model based diagnosis methodologies for pem fuel cell stacks and systems," *International Journal of Hydrogen Energy*, vol. 38, no. 21, pp. 8914–8926, 2013.

[57] J. Hua, L. Lu, M. Ouyang, J. Li, and L. Xu, "Proton exchange membrane fuel cell system diagnosis based on the signed directed graph method," *Journal of Power Sources*, vol. 196, no. 14, pp. 5881–5888, 2011.

[58] P. M. Frank, "Non-analytical approaches to model-based fault detection and isolation," *Advances in Automatic Control*, vol. 754, pp. 85–99, 2004.

[59] D. Gao, C. Wu, B. Zhang, and X. Ma, "Signed directed graph and qualitative trend analysis based fault diagnosis in chemical industry," *Chinese Journal of Chemical Engineering*, vol. 18, no. 2, pp. 265–276, 2010.

[60] B. Ould-Bouamama, R. El Harabi, M. N. Abdelkrim, and M. K. Ben Gayed, "Bond graphs for the diagnosis of chemical processes," *Computers & Chemical Engineering*, vol. 36, no. 10, pp. 301–324, 2012.

[61] Y. H. Wong, A. B. Rad, and Y. K. Wong, "Qualitative modeling and control of dynamic systems," *Engineering Applications of Artificial Intelligence*, vol. 10, no. 5, pp. 429–439, 1997.

[62] R. Smaili, R. El Harabi, and M. N. Abdelkrim, "FDI based on causal graphical approaches for nonlinear processes," *10th International Multi-Conference on Systems, Signals & Devices (SSD)*, pp. 1–6, 2013.

[63] C. H. Lo, Y. K. Wong, and A. B. Rad, "Intelligent system for process supervision and fault diagnosis in dynamic physical systems," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 2, pp. 581–592, 2006.

[64] V. Venkatasubramanian, R. Raghunathan, and N. K. Surya, "A review of process fault detection and diagnosis part II: Qualitative models and search strategies," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 313–326, 2003.

[65] J. Zhang, P. D. Roberts, and J. E. Ellis, "Fault diagnosis of a mixing process using deep qualitative knowledge representation of physical behaviour," *Journal of Intelligent and Robotic Systems*, vol. 3, no. 2, pp. 103–115, 1990.

[66] J. Montmain and S. Gentil, "Dynamic causal model diagnostic reasoning for online technical process supervision," *Automatica*, vol. 36, no. 8, pp. 1137–1152, 2000.

[67] A. H. Ali, D. Dubois, and H. Prade, "Qualitative reasoning based on fuzzy relative orders of magnitude," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 1, pp. 9–23, 2003.

[68] I. B. Ozyurt, L. O. Hall, and A. K. Sunol, "Sqfdiag: semi-quantitative model-based fault monitoring and diagnosis via episodic fuzzy rules," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 29, no. 3, pp. 294 –306, 1999.

[69] A. Evsukoff, S. Gentil, and J. Montmain, "Fuzzy reasoning in co-operative supervision systems," *Control Engineering Practice*, vol. 8, no. 4, pp. 389–407, 2000.

[70] J. Rasmussen, *Information processing and human-machine interaction*. New York: North Holland, 1986.

[71] J. Arenas-Garcia, K. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 16–29, 2013.

[72] B. K. Lavine and W. S. Rayens, "Statistical discriminant analysis," *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, from Comprehensive Chemometrics*, pp. 517–540, 2012.

[73] J. Jackson, *A User's Guide to Principal Components*, ser. Wiley Series in Probability and Statistics. New York: Wiley, 2003.

[74] I. Jolliffe, "Chapter three - properties of sample principal components," in *Principal Component Analysis (Second Edition)*, 2002, pp. 265–304.

[75] T.-N. Yang and S.-D. Wang, "Robust algorithms for principal component analysis," *Pattern Recognition Letters*, vol. 20, no. 9, pp. 927–933, 1999.

[76] S. Serneels and T. Verdonck, "Principal component analysis for data containing outliers and missing elements," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1712–1727, 2008.

[77] S. W. Choi, C. Lee, J. M. Lee, J. H. Park, and I. B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 1, pp. 55–67, 2005.

[78] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1960–1971, 2011.

[79] J. M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 59, no. 1, pp. 223–234, 2004.

[80] R. Luo, M. Misra, and D. M. Himmelblau, "Sensor fault detection via multiscale analysis and dynamic pca," *Industrial & Engineering Chemistry Research*, vol. 38, no. 4, pp. 1489–1495, 1999.

[81] C. Lee, S. W. Choi, and I.-B. Lee, "Sensor fault identification based on time-lagged PCA in dynamic processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 70, no. 2, pp. 165–178, 2004.

[82] W. Li, H. H. Yue, S. Valle-Cervantes, and S. J. Qin, "Recursive PCA for adaptive process monitoring," *Journal of Process Control*, vol. 10, no. 5, pp. 471–486, 2000.

[83] J. Tang, W. Yu, T. Chai, and L. Zhao, "On-line principal component analysis with application to process modeling," *Neurocomputing*, vol. 82, pp. 167–178, 2012.

[84] J. L. Godoy, J. R. Vega, and J. L. Marchetti, "A fault detection and diagnosis technique for multivariate processes using a PLS-decomposition of the measurement space," *Chemometrics and Intelligent Laboratory Systems*, vol. 128, pp. 25–36, 2013.

[85] S. Wold, M. Sjostrom, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.

[86] D. Zhou, G. Li, and S. J. Qin, "Total projection to latent structures for process monitoring," *AIChE Journal*, vol. 56, no. 1, pp. 168–178, 2010.

[87] G. Li, B. Liu, S. J. Qin, and D. Zhou, "Quality relevant data-driven modeling and monitoring of multivariate dynamic processes: The dynamic t-pls approach," *IEEE Transactions on Neural Networks*, vol. 2, no. 12, pp. 2262–2271, 20.

[88] S. J. Qin and Y. Zheng, "Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures," *AIChE Journal*, vol. 59, no. 2, pp. 496–504, 2013.

[89] Q. Liu, S. Qin, and T. Chai, "Multiblock concurrent PLS for decentralized monitoring of continuous annealing processes," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6429–6437, 2014.

[90] K. Kim, J.-M. Lee, and I.-B. Lee, "A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 79, no. 1-2, pp. 22–30, 2005.

[91] S. M. Namburu, M. S. Azam, J. Luo, K. Choi, and K. R. Pattipati, "Data-driven modeling, fault diagnosis and optimal sensor selection for hvac chillers," *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 3, pp. 469–473, 2007.

[92] M. Delgado, A. Garcia, J. A. Ortega, J. J. Cardenas, and L. Romeral, "Multidimensional intelligent diagnosis system based on support vector machine classifier," *IEEE International Symposium on Industrial Electronics (ISIE)*, pp. 2124–2131, 2011.

[93] C. Sankavaram, B. Pattipati, K. Pattipati, Y. Zhang, M. Howell, and M. Salman, "Data-driven fault diagnosis in a hybrid electric vehicle regenerative braking system," *IEEE Aerospace Conference*, pp. 1–11, 2012.

[94] D.-T. Pham, "Fast algorithms for mutual information based independent component analysis," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2690–2700, 2004.

[95] D. Erdogmus, K. Hild, Y. Rao, and J. Principe, "Minimax mutual information approach for independent component analysis," *IEEE Transactions on Signal Processing*, vol. 16, no. 6, pp. 1235–1252, 2004.

[96] C. Jingyan, Y. Jie, J. Mori, M. M. Rashid, H. Gangshi, Y. Honglu, J. Flores-Cerrillo, and L. Megan, "An independent component analysis and mutual information based non-gaussian pattern matching method for fault detection and diagnosis of complex cryogenic air separation process," *American Control Conference (ACC)*, pp. 2797–2802, 2013.

[97] G. Stefatos and A. B. Hamza, "Dynamic independent component analysis approach for fault detection and diagnosis," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8606–8617, 2010.

[98] Y. Zhang, J. An, and C. Ma, "Fault detection of non-gaussian processes based on model migration," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 5, pp. 1517–1526, 2007.

[99] Y. Guo, J. Na, B. Li, and R.-F. Fung, "Envelope extraction based dimension reduction for independent component analysis in fault diagnosis of rolling element bearing," *Journal of Sound and Vibration*, vol. 333, no. 13, pp. 2983–2994, 2014.

[100] Z. Wang and C. S. Chang, "Online fault detection of induction motors using frequency domain independent components analysis," *IEEE International Symposium on Industrial Electronics (ISIE)*, pp. 2132–2137, 2011.

[101] A. Widodo, B.-S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors," *Expert Systems with Applications*, vol. 32, no. 2, pp. 299–312, 2007.

[102] Q. He, Z. Feng, and F. Kong, "Detection of signal transients using independent component analysis and its application in gearbox condition monitoring," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2056–2071, 2007.

[103] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measurement*, vol. 46, no. 1, pp. 259–271, 2013.

[104] H. C. Dubey, S. R. Mohanty, and N. Kishore, "Abrupt change detection of fault in power system using independent component analysis," *International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN)*, pp. 659–664, 2011.

[105] A. Ajami and M. Daneshvar, "Data driven approach for fault detection and diagnosis of turbine in thermal power plant using independent component analysis (ICA)," *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 728–735, 2012.

[106] M. Ruiz-Llata, G. Guarnizo, and C. Boya, "Embedded power quality monitoring system based on Independent Component Analysis and SVMs," *The 2011 International Joint Conference on Neural Networks (IJCNN)*, pp. 2229–2234, 2011.

[107] A. Lebaroud and G. Clerc, "Classification of induction machine faults by optimal time-frequency representations," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4290– 4298, 2008.

[108] C. Delpha, D. Diallo, M. 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.

[109] J. Ye, "Characterization of a family of algorithms for generalized discriminant analysis on under sampled problems," *Journal of Machine Learning Research*, vol. 6, pp. 483–502, 2005.

[110] Z. Zhao and T. W. S. Chow, "Tensor locally linear discriminative analysis," *IEEE Signal Processing Letters*, vol. 18, no. 11, pp. 643–646, 2011.

[111] ——, "Robust linearly optimized discriminant analysis," *Neurocomputing*, vol. 79, pp. 140–157, 2012.

[112] J. Li and P. Cu, "Improved kernel fisher discriminant analysis for fault diagnosis," *Expert Systems with Applications*, vol. 36, no. 1, pp. 1423–1432, 2009.

[113] M. Fernandez-Temprano, P. E. Gardel-Sotomayor, O. Duque-Perez, and D. Morinigo-Sotelo, "Broken bar condition monitoring of an induction motor under different supplies using a linear discriminant analysis," *9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED)*, pp. 162–168, 2013.

[114] M. Sahni and W. J. Lee, "Classification of severity of low-voltage motor coil arcing fault using statistical techniques," *IET Generation, Transmission & Distribution*, vol. 3, no. 1, pp. 75–85, 2009.

[115] R. Z. Haddad and E. G. Strangas, "Fault detection and classification in permanent magnet synchronous machines using fast fourier transform and linear discriminant analysis," *9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED)*, pp. 99–104, 2013.

[116] W. Li, L. Zhang, and Y. Xu, "Gearbox pitting detection using linear discriminant analysis and distance preserving self-organizing map," *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 2225–2229, 2012.

[117] J. Xiaohang, Z. Mingbo, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014.

[118] Y. Yang, Y. Chen, X. Chen, and X. Liu, "Multivariate industrial process monitoring based on the integration method of canonical variate analysis and independent component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 116, pp. 94–101, 2012.

[119] H. Borsje, "Fault detection in boilers using canonical variate analysis," *Proceedings of the 1999 American Control Conference*, vol. 2, pp. 1167–1170, 1999.

[120] R. Zimroz and A. Bartkowiak, "Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 237–247, 2013.

[121] S. Hamdani, A. Bouzida, O. Touhami, and R. Ibtiouen, "Diagnosis of rotor fault in induction motor using the music analysis of the terminal voltage after

switch-off," *18th International Conference on Electrical Machines (ICEM)*, pp. 1–5, 2008.

[122] A. Garcia-Perez, R. de Jesus Romero-Troncoso, E. Cabal-Yepez, and R. A. Osornio-Rios, "The application of high-resolution spectral analysis for identifying multiple combined faults in induction motors," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 2002–2010, 2011.

[123] B. Kilundu, P. Dehombreux, and X. Chiementin, "Tool wear monitoring by machine learning techniques and singular spectrum analysis," *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 400–415, 2011.

[124] R. Solimene and G. Leone, "Music algorithms for grid diagnostics," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 2, pp. 226–230, 2013.

[125] B. Xu, L. Sun, L. Xu, and G. Xu, "Improvement of the Hilbert method via ESPRIT for detecting rotor fault in induction motors at low slip," *IEEE Transactions on Energy Conversion*, vol. 28, no. 1, pp. 225–233, 2013.

[126] B. Muruganatham, M. A. Sanjith, B. Krishnakumar, and S. A. V. Satya Murty, "Roller element bearing fault diagnosis using singular spectrum analysis," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 150–166, 2013.

Chapter 2

Kullback-Leibler Divergence for Fault Detection

2.1 Introduction

Measurements are basic representations of process behaviour, and faults in general manifest themselves as changes in their properties. Detecting the particular fault that occurs in a system is based on checking whether the current measurements are statistically different from the *a priori* known faultless measurements. Detection indicators with adequate thresholds are designed to this end. They may operate either directly on the measurements acquired from the process, or on their transformations (residuals, features, components, etc.) obtained with the modelling approaches discussed in chapter 1. There are two main approaches for designing appropriate fault indicators, the statistical hypothesis testing approach and the norm-based approach, see [1] and the references therein. The statistical hypothesis testing approach is concerned with detecting a change in the probability distribution of a process. It results in three main types of statistical control charts, namely the Shewhart [2], the cumulative sum (CUSUM) [3] and the exponentially weighted moving average (EWMA) [4]. The norm-based approach uses a varying or constant threshold on some norm of the measurements (residuals or components).

The statistical literature reports two classes of control charts: the distribution-based (parametric) [5, 6, 7] and the distribution-free (nonparametric) [8, 9, 10] control charts. The distribution-based control charts assume *a priori* known probability distribution (usually normal for continuous measurements) and aim at detecting shifts in one of its parameters (mean, scale, kurtosis, etc.). The distribution-free control charts make no assumption for the form of the process distribution, but require the availability of a training sample of observations from which the reference

(fault-free) empirical probability distribution can be calculated. Their advantage over the distribution-based is that a general rather than a one-parameter change in the process distribution can be monitored [11]. Despite this fact, most attention in both theoretical and practical studies has been paid to the parametric techniques, which definitely become the traditional statistical techniques applied to process monitoring.

Kullback-Leibler (KL) divergence is a measure of dissimilarity between two probability distributions and plays a key role in solving information theory and change detection problems. The literature on change detection and FDD reports interesting theoretical findings about the contribution of the divergence properties to deriving fault detection and estimation indices [12, 13, 14]. To the best of our knowledge, these studies are notably developed in a parametric model-based framework, whose main concern is to detect a particular change in model parameters of a process or signal. This chapter seeks to answer these questions: Does the KL divergence itself can be used as a fault indicator, into rather a general, distribution-free and non-parametric framework? Is it sensitive to incipient faults? How does the fault detection with KL divergence perform in noisy environments?

So, this chapter addresses the problem of detecting incipient faults which cause unpredictable changes, near the noise level, in the process data. More details on the proposal motivations are given in the sequel.

2.2 Motivations and Outline

Providing that assumptions regarding the properties (form and parameters) of the process probability distribution, and the properties of the faults to be detected including how they may affect the monitored system, can be made, the conventional parametric statistical control charts are very effective in identifying the considered faults and statistically monitoring the system [15, 16, 17]. This prior knowledge may however not be available in practice. The process probability distribution may have an arbitrary form which is far from being normal. The fault may affect several parameters, or the information about its impact on the measurements may be lacking. If the fault is serious enough, it causes a large change easily detectable with the usual parametric techniques even if the underlying assumptions are not respected. However, the critical case is when the abnormal change following the fault occurrence is unpredictable and maybe concealed by other sources (normal variations, noises, etc.), typically because the fault is incipient, or the noise level is high enough so as to corrupt the detection of the incipient deviation. The conventional control charts are less efficient in this case. To cope with such situations, a non-parametric fault

indicator requiring little *a priori* knowledge and aiming at detecting global changes is needed.

Assuming that the process probability distribution is stationary, i.e. time-independent, Kullback-Leibler divergence, also called relative entropy, is proposed in this work to globally monitor the statistical process behaviour. As a distribution-free fault indicator, the divergence quantifies explicitly the general discrepancy or difference between the current distribution calculated using the current observation set and the reference one calculated from a training observation set. The KL divergence is an instance of f-divergence family which has been used in many signal processing applications including anomaly and change detection [18, 19], pattern recognition [20, 21], classification and coding [22, 23]. It however has a particular place in the model-based FDD framework compared to the other f-divergence criteria. It has been used to derive discrimination indices between parametric system/signal models for fault detection purposes, see [12]. Theoretical studies in the context of abrupt change detection had led to the KL divergence as a measure of detectability of abrupt changes in parameters of system/signal models [24, 13]. Assuming normality distribution of model-based residuals allowed designing statistical control charts for change detection. For fault estimation, the divergence between AR/ARMA models and between state space models before and after abrupt changes was expressed in function of the generated residuals (innovations).

The KL divergence between arbitrary probability distributions has no closed form. Its use in model-based FDD is based on the assumption of normal probability distribution leading to a closed divergence expression. Here we intend to evaluate the ability to detect faults without considering a specific distribution, but by using a numerical estimation of the divergence. Besides, it is generally always useful to reduce the dimensionality of the process data that involve several related variables. This is achievable by decorrelating the variables and extracting data components with high variability. PCA provides these features and asks no particular information about the process. It is optimal in terms of capturing variability in the data, and constitutes a general framework for data representation and modelling. It has been used for monitoring in a wide range of applications, including chemical processes, water treatement [25], manufacturing, aerospace [26, 27], electronics [28], automotive [29], semi-conductors [30], and many others. Therefore, it seems meaningful to adopt rather a PCA-based Kullback Leibler divergence technique for the statistical process monitoring.

The divergence will be applied, without restriction, to the data principal components. Its efficiency in detecting incipient faults will be compared with that of

usual statistics, namely the Hotelling $\mathbf{T^2}$ and the squared prediction error (SPE). An incipient fault is often defined as a change or a degradation that develops slowly. The fault model adopted here assumes that during the first stage of the incipient fault development, the fault amplitude (size or severity) is constant, see Fig.2.1. Obviously, the difficulty is to detect the incipient fault in this stage where a signif-

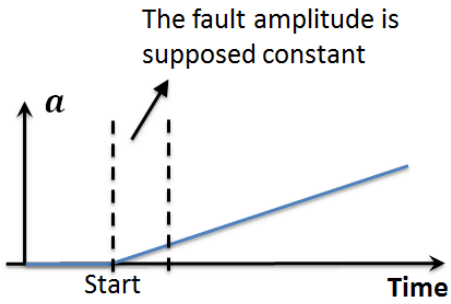

Figure 2.1: Incipient fault model

icant loss of process operation performance can occur without being noticed. The detection error rates will be calculated in order to evaluate the detection performance of the divergence in real-world noisy environments. We refer to the Fault-to-Noise Ratio (FNR) as a comparative criterion between the fault and noise levels. The region around 0 dB of FNR is particularly of interest, since it refers to a critical situation for fault detection: the change following the fault occurrence is submerged in noise. Note that with the assumption of incipient fault detection, the fault level is very small compared to the original signal level.

In the following, our proposal concerning a PCA-based KL divergence for fault detection will be detailed. Section 3 is devoted to the KL divergence definition, estimation and role in model-based FDD. Section 3 describes the procedure of detection performance evaluation using detection error probabilities. Section 4 is dedicated to PCA modelling and fault detection. Section 5 is concerned with the validation, through simulation, of our proposal.

2.3 Kullback-Leibler divergence

2.3.1 Definition

The Kullback-Leibler divergence, or the relative entropy, is a widely used probabilistic tool in telecommunications and multimedia security. It plays a key role in machine learning and neuroscience. It was successfully applied for pattern recognition [31, 32], anomaly detection [33, 34, 35], and classification [36, 37]. It also has a

particular place in model-based FDD.

For computing dissimilarity between two probability density functions (pdfs) $h(z)$ and $q(z)$ of a continuous random variable z, Kullback and Leibler define the Kullback-Leibler Information from h to q as [38]

$$I(h\|q) = \int h(z) \log \frac{h(z)}{q(z)} dz \tag{2.1}$$

$$= \mathbb{E}_h\{log\frac{h(z)}{q(z)}\} \tag{2.2}$$

where $\mathbb{E}_h$ means the expectation over the distribution h. The divergence is then defined as the symmetric version of the Information and is given by

$$D(h, q) = I(h\|q) + I(q\|h). \tag{2.3}$$

It is non-negative and null if and only if the two distributions are strictly equal. It is not a true metric (does not satisfy the triangular inegality) and is only defined if $\int h dz = 1$, $\int q dz = 1$ and if the two distributions share the same support set.

2.3.2 KL divergence in model-based change detection

The KL divergence has a particular place in model-based FDD, due to its strong connection with the logarithm of the likelihood ratio. The log-likelihood ratio is itself a key concept in mathematical statistics and is the basis theory in designing statistical decision functions between 'no-fault' and 'fault' hypotheses. Statistical decision techniques, commonly Schewhart and Moving average control charts, CUSUM and GLR (Generalised Likelihood Ratio) algorithms, were designed to detect abrupt changes in signal characteristics based on log-likelihood ratio properties. The basic types of changes are deviation from the reference mean value and scale increase in the observed signals laws.

Let's assume given a finite observation set z_1,z_2,...,z_N , and two possible models $\mathcal{H}_0$ and $\mathcal{H}_1$ which refer to the null hypothesis 'No-fault (without changes)' and the alternative hypothesis 'Fault (with changes)' respectively. Suppose afterwards that the sample has a probability density or a mass function h_θ parametrised with θ, such that $\theta = \theta_0$ under $\mathcal{H}_0$, and $\theta = \theta_1$ under $\mathcal{H}_1$. Then the log-likelihood ratio is:

$$L(z) = \log \frac{h_{\theta_1}(z_1, z_2, ..., z_N)}{h_{\theta_0}(z_1, z_2, ..., z_N)} \tag{2.4}$$

The joint distribution is equal to the product of marginal distributions if the observations are independently distributed. In this case $L(z)$ is:

$$L(z) = \sum_{i=1}^{N} \log \frac{h_{\theta_1}(z_i)}{h_{\theta_0}(z_i)} \tag{2.5}$$

The larger $L(z)$ is, more evidence exists in favour of the alternative hypothesis. Given θ_0, the change from θ_0 to θ_1 is defined by θ_1 that maximises $L(z)$. Intuitively, it follows that

$$\mathbb{E}_{h_{\theta_1}}(L(z)) > \mathbb{E}_{h_{\theta_0}}(L(z)). \tag{2.6}$$

where $\mathbb{E}_{\theta_1}\{L(z)\}$ is the expectation of $L(z)$ if the 'fault (with change)' hypothesis is true, and $\mathbb{E}_{\theta_0}\{L(z)\}$ is the expectation of $L(z)$ if the 'no-fault (without change)' hypothesis is true. This corresponds to the statistical detectability definition: a change is said to be statistically detectable if the mean value of the log-likelihood ratio is higher after change than before change. In terms of KL divergence, this is equivalent to:

$$D(h_{\theta_1}, h_{\theta_0}) = \mathbb{E}_{h_{\theta_1}}(L(z)) - \mathbb{E}_{h_{\theta_0}}(L(z)) > 0 \tag{2.7}$$

The divergence value is subsequently a measure of detectability of an abrupt change. Since the detectability increases with the change magnitude as well as the signal-to-noise ratio (SNR), it follows that the divergence can likewise be viewed as a measure of the change magnitude. These intuitive concepts are developed in [13], in a more formal theoretical study.

2.3.3 Estimation

We assume no specific parametric model for observed signals or system. For fault detection purposes, we suppose that little information about the change's type and the fault model is available. We intend to evaluating in these conditions the ability of the divergence to detect small changes due to fault occurrence.

As the divergence between two arbitrary probability distributions has no closed form, the integral function given by Eq.(2.3) should be numerically approximated. The common method to estimate the divergence value uses the interpretation of the Information in term of the likelihood ratio: the KL information from probability distribution h to q is the expected log-likelihood ratio $\log(h/q)$ under the distribution h. This induces two assumptions:

1. an observation set composed of N independent and identically distributed (i.i.d.) observations $\{z_i\}_1^N$ drawn from h is supposed available.

2. $q(z_i)$ can be calculated, and thus q is supposed to be known.

Under these assumptions, the Monte Carlo approximation consists in computing:

$$I_{MC}(h\|q) = \frac{1}{N}\sum_{i=1}^{N}\log\frac{h(z_i)}{q(z_i)} \tag{2.8}$$

The law of large numbers asserts that the variable $\sqrt{N}(I_{MC}(h\|q) - I(h\|q))$ is approximately normal with zero mean and variance $\sigma^2_{MC} = Var_h\left[log(h/q)\right]$ as $N \to \infty$. In our application, the probability distributions are not known *a priori* and for detection, we do not assume any specific form. However, two observation sets are available, the current and the reference, from which empirical probability density functions can be calculated using kernel density estimators for example.

An intuitive and fast way to approximate the divergence between two unknown probability distributions consists in the discrete form of the divergence that uses empirical probability density functions. Kernel density estimators are non-parametric estimators whose parameters depend on the data. The Gaussian kernel is often used due to its mathematical properties and to the law of large numbers that makes distributions converge to normal as the sample size grows. Consider a partition of the reference observation set into n_I disjoint intervals $\{[s_0, s_0 + \Delta s), ..., [s_0 + (n_I - 1)\Delta s, s_{n_I})\}$ where s_0 and s_{n_I} are the min and max observation values. The probabilities $\{\zeta_1,\zeta_2,...,\zeta_{n_I}\}$ of the reference observation values are estimated by applying a Gaussian kernel function to each interval centre. The probabilities $\{\nu_1, \nu_2, ..., \nu_{n_I}\}$ of the current observation values are calculated similarly for the same set of intervals. Then the divergence is approximated by:

$$\hat{D} = \sum_{i=1}^{n_I}(\zeta_i - \nu_i)\log\frac{\zeta_i}{\nu_i} \tag{2.9}$$

Theoretically if the current pdf is equal to the reference one, the divergence will be null. Otherwise, a fault or an abnormal process event causes some change that appears as a disparity between current and reference pdfs. This disparity is measured by $\hat{D}$. Subsequently, the divergence value is likely to represent the fault severity: the more severe the fault is, the greater the distance from the distribution to its reference. In practice however, there is always a low non-zero divergence caused by random noise from real data and numerical errors. Therefore, there is no guarantee for the divergence to take a null value in no-fault condition. So the decision whether the system has entered into abnormal operating conditions must be made with respect to a threshold that we refer to as ϵ_{safe}. The no-fault condition becomes:

$$\hat{D} \leq \epsilon_{safe}. \tag{2.10}$$

ϵ_{safe} is the nominal value of $\hat{D}$ calculated from training data that are recorded

under normal operating conditions. Constant threshold requires that the probability distributions are stationary. This is plausible in three ways:

- The data used to estimate the probability distributions don't exhibit non-stationary features, such as non-stationary harmonics and time-dependent noise.
- The data size is sufficiently large so as to mitigate the effect of non-stationary features.
- The data can be preprocessed to remove the non-stationary features based on the prior knowledge of these features.

2.4 Detection performance evaluation

The fault detection that uses the divergence as a decision function relies on the estimation of the divergence value (Eq.2.9). This value is compared to a predetermined threshold ϵ_{safe} to make decision. The threshold should be set for a particular application according to the performance specifications required. False alarm and missed alarm rates are the most important detection performance criteria, from a practical point of view.

The fault detection is then based on the following hypothesis test:

$$\hat{D} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \epsilon_{safe} \tag{2.11}$$

where $\mathcal{H}_0$ is the 'no-fault' hypothesis and $\mathcal{H}_1$ is the 'fault' hypothesis. The performance of the test is characterised with probabilities of false alarm (P_{FA}) and missed detection (P_{MD}):

$$\begin{aligned} P_{FA} &= P(\hat{D} > \epsilon_{safe}|\mathcal{H}_0) \\ P_{MD} &= P(\hat{D} < \epsilon_{safe}|\mathcal{H}_1). \end{aligned} \tag{2.12}$$

The divergence $\hat{D}$ calculated under specific conditions of noise and fault can be assumed as a random variable that has a specific distribution with a parameter vector θ. Consequently, the error probabilities become:

$$\begin{aligned} P_{FA} &= P(\hat{D} > \epsilon_{safe}|\theta_0) \\ P_{MD} &= P(\hat{D} < \epsilon_{safe}|\theta_1). \end{aligned} \tag{2.13}$$

For a given value of P_{FA}, the probability of missed detection allows evaluating the sensitivity of the divergence to small faults in the presence of noise. So let's introduce the fault-to-noise ratio (FNR) as a comparative criterion between the noise and fault levels. It is given by:

$$FNR = 10 * log_{10} \frac{\sigma_f^2}{\sigma_v^2} \tag{2.14}$$

where σ_f^2 is the power of the change following the fault occurrence, and σ_v^2 is the noise power. In practice, one should have prior knowledge on the noise and the profile of the fault in order to calculate the experimental FNR. Incipient faults correspond to high values of signal-to-fault ratio (SFR). Their detection becomes challenging when the fault level ranges close to the noise level, inducing near-zero FNR values. SFR and FNR are linked to the SNR by the following equation:

$$SNR = SFR + FNR \tag{2.15}$$

It stems from Eq.2.15 that dealing with incipient faults (high SFR) that are masked by noise (near-zero FNR) means that the SNR is positive. In other words, having both a low SNR and a low FNR induces a low SFR which means that the fault is no longer incipient: the change caused by the fault is not small with respect to the signal.

A noisy reference probability distribution corrupts the sensitivity of the fault indicator, here the divergence, with respect to incipient faults. This can be overcome by several means, including the following:

- The noise effect can be notably reduced by estimating the probability distributions on large observation sets (theoretically $N \to \infty$).
- The noise can be filtered out assuming that *a priori* knowledge about its nature and its main characteristics are available.
- In worst case, a noisy reference will result in a higher threshold value ϵ_{safe} for the divergence, thus reducing the sensitivity to small faults.

The sensitivity of KL divergence with respect to incipient faults that are masked by noise will be evaluated in this chapter on a numerical example. Before going through the results, the fault detection using PCA will be briefly reviewed, and then the proposed divergence application will be described.

2.5 PCA for fault detection and diagnosis

The fault detection and diagnosis using PCA has received considerable attention in the last two decades. The advantage of PCA over other fault diagnosis approaches (chapter 1) is that the development of the reference model to which the operating process will be compared to generate decision functions about the system health state, requires no complicated physical knowledge on the process, neither detailed theoretical studies to be available. The only information needed is a good historical database describing the normal process operation. The basic idea of PCA has been explained in Chapter 1, section 1.3.2. We also reviewed some of the most interesting PCA variants that were proposed in the literature to cope with the limitations of the classical linear version. We consider here, without restriction, the application of the KL divergence to the principal components generated with the classical PCA technique.

2.5.1 PCA's model identification

Let us consider N measurements/observations of m correlated variables $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, ..., $\boldsymbol{x_m}$ that are collected at different sampling instants i from the process. They are arranged into columns of a matrix $X_{[N\times m]}$, $X = \left[\ x(1),\ \cdots,\ x(i),\ \cdots,\ x(N)\ \right]'$, where $x(i) \in \mathbb{R}^m$ is the ith measurement of the m variables. The data are recorded when the process is in state of control (healthy operating mode) so that the PCA's model that will be built acts as a reference of the normal process behaviour. The analysis can be made on either the covariance or the correlation data matrix [39]. Although, it is often wise to work with a correlation matrix which normalizes the variances of the variables before applying PCA. So the different underlying relationships between variables could be extracted. The covariance matrix has however advantages over the correlation, especially in cases where PCA is used as an inferential and not only descriptive tool. Statistical inferences about principal components population are easier to derive for the covariance matrix than for the correlation one [40]. So, let $\bar{X}$ denote the autoscaled matrix of X. The columns of $\bar{X}$ are either centred or centred and reduced, depending on whether the covariance or the correlation matrix is used. PCA transforms the data matrix $\bar{X}$ into a new matrix $T_{[N\times m]}$ of uncorrelated variables $\boldsymbol{t_1}, \boldsymbol{t_2}, ..., \boldsymbol{t_m}$ termed principal component scores, $T = \left[\ t(1),\ \cdots,\ t(i),\ \cdots,\ t(N)\ \right]'$, where $t(i) \in \mathbb{R}^m$. The new variables are linear combinations of the original ones and they successively maximise the total data variance. T is obtained from $\bar{X}$ by an orthogonal transformation whose vectors correspond to the eigenvectors of the covariance/correlation matrix. These vectors, termed the loading vectors, are arranged as columns of a matrix $P_{[m\times m]}$ in the descendent order of their corresponding eigenvalues.

According to these elements, the solution to the PCA problem given by the eigenvector decomposition of the covariance/correlation matrix is:

$$S = \frac{1}{N}\bar{X}'\bar{X} = P\Lambda P' \quad \text{with} \quad PP' = P'P = I_m \tag{2.16}$$

where $\Lambda = diag(\lambda_1, ..., \lambda_m)$ is a diagonal matrix containing the eigenvalues of S in a descendent order, I_m is the identity matrix, and prime denotes the transpose operator. The scores matrix and the original data matrix are then:

$$T_{[N\times m]} = \bar{X}_{[N\times m]} P_{[m\times m]} \tag{2.17}$$

$$\bar{X}_{[N\times m]} = T_{[N\times m]} P'_{[m\times m]} \tag{2.18}$$

The data dimensionality reduction is achieved by splitting $P - \Lambda$ into two parts, $\hat{P}_{[m\times l]} - \hat{\Lambda}_l$ and $\tilde{P}_{[m\times(m-l)]} - \tilde{\Lambda}_{m-l}$ spanning respectively the principal and the residual subspaces. $\hat{P}_{[m\times l]}$ contains the first l loading vectors associated with the largest eigenvalues $\hat{\Lambda}_l$, and $\tilde{P}_{[m\times(m-l)]}$ contains the last vectors. Accordingly, the first l columns $\boldsymbol{t_1}, \boldsymbol{t_2}, ..., \boldsymbol{t_l}$ of the scores matrix T represent the important variability of data, and the last $(m - l)$ ones, $\boldsymbol{t_{l+1}}, ..., \boldsymbol{t_m}$, represent all forms of residues including noise and redundancy. The residual principal components identify quasi-linear constant relationships between the original variables.

l is subsequently the new data dimension. $\bar{X}$ can be approximated by its projection into the principal subspace as:

$$\widehat{X} = \bar{X}\hat{P}\hat{P}' \tag{2.19}$$

with an approximation error given by

$$\tilde{X} = \bar{X} - \widehat{X} = \bar{X}(I - \hat{P}\hat{P}'). \tag{2.20}$$

$\tilde{X}$ is effectively the residual of the data, that is its projection into the residual space.

The $[m \times m]$ matrix $\widehat{C} = \hat{P}\hat{P}'$ can be thus considered as an explicit PCA model. It only depends on l which is the parameter to be identified. Several studies have been conducted to show the relationship between the fault detection ability and the number of principal components retained in the PCA model, see for example [41, 42]. Many criteria have been proposed in the literature to get the best choice l, such as the cumulative percentage of total variation [43] and the cross-validatory criterion [44]. Furthermore, some more interesting criteria are devoted for fault detection and identification purposes such as minimizing the Variance of Reconstruction Error

(VRE) [45]. The reconstruction consists in estimating each variable using the PCA model that is a function of l and the remaining variables. The optimal l is the number minimising the reconstruction error, which is the difference between the variable and its reconstruction. Authors in [46] compared 11 methods to determine l and concluded that the VRE criterion is preferable.

2.5.2 PCA-based fault detection

2.5.2.1 Distance-based statistics

Typical detection indices are distance-based, aiming at evaluating how much a new observation is away from each of the subspaces [47]. The fault detection is based on a PCA model that has been built from fault-free training data. Each new observation vector $x(i)$ taken at time instant i is decomposed using the model into two orthogonal parts, approximate vector $\hat{x}(i)$ and residual vector $\tilde{x}(i)$,

$$\widehat{x}(i) = \bar{x}(i)'\widehat{C} \quad \text{and} \quad \tilde{x}(i) = \bar{x}(i)'(I - \widehat{C}). \tag{2.21}$$

Both parts are monitored using detection indices. The approximate part identifies outliers in the majority of data. The residual part allows monitoring the data correlation structure. Hotelling $\mathbf{T^2}$ and SPE statistics are commonly applied for this purpose. Sensor faults are particularly concerned by this monitoring approach. Given that the scores vector of $x(i)$ is $t(i) = \bar{x}(i)'P$ such that $t(i) = [t_{i1},\ ...,\ t_{il},\ ...,\ t_{im}]$, the $\mathbf{T^2}$ and the SPE at instant i are:

$$\begin{aligned} \mathbf{T^2}(i) &= \hat{x}(i)'\hat{P}\hat{\Lambda}\hat{P}'\hat{x}(i) && (2.22)\\ &= \sum_{k=1}^{l} \frac{t_{ik}^2}{\lambda_k} && (2.23) \end{aligned}$$

$$\begin{aligned} SPE(i) &= \|\tilde{x}(i)\| && (2.24)\\ &= \sum_{k=l+1}^{m} t_{ik}^2 && (2.25) \end{aligned}$$

$x(i)$ is declared out-of-control (faulty) if

$$\mathbf{T^2}(i) > \mathbf{T^2}_{l,\alpha} \quad \text{and/or} \quad SPE(i) > \delta_\alpha^2 \tag{2.26}$$

where $\mathbf{T^2}_{l,\alpha}$ and δ_α^2 are the control limits or the thresholds at significance level α. The distance-based fault detection is summarised in Fig.2.2.

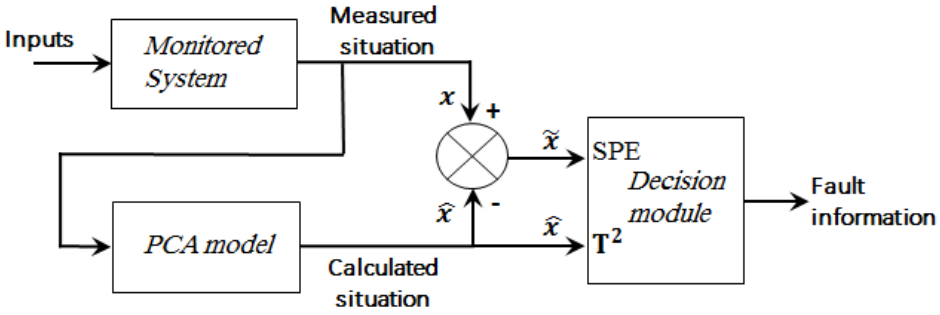

Figure 2.2: Distance-based fault detection

Theoretical thresholds are developed for large sample size N and independent multivariate normally distributed data. The thresholds calculated directly from training data are however often more effective in practice. Box [48] gives a theoretical threshold for SPE, assuming the data have multivariate normal distribution. Suppose $\gamma_c = \sum_{k=l+1}^{m} \lambda_k^c$ where λ_k^c is the kth eigenvalue to the cth power. The theoretical threshold of SPE at significance level α is:

$$\delta_\alpha^2 = g\chi_{h,\alpha}^2 \tag{2.27}$$

where $g = \gamma_2/\gamma_1$, $h = integer(\gamma_1^2/\gamma_2)$, $integer(o)$ is the integer value of o and $\chi_{h,\alpha}^2$ is the Chi- square distribution with h degrees of freedom.

A threshold for $\mathbf{T^2}$ can be approximated by:

$$\mathbf{T^2}_{l,\alpha} = \frac{l(N^2-1)}{N(N-l)} F_{l,N-l,\alpha} \tag{2.28}$$

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

The sensitivity of SPE and $\mathbf{T^2}$ to the faults depends on the chosen l. Variants of SPE, such as the squared weighted error SWE, were proposed to reduce the effect of l setting on the detection performance [49]. The SWE is:

$$SWE(i) = \sum_{k=l}^{m+1} \frac{t_{ik}^2}{\lambda_k} \tag{2.29}$$

Some combined statistics were also proposed, as for instance [50]:

$$\phi(i) = \frac{SPE(i)}{\delta_\alpha^2} + \frac{\mathbf{T^2}(i)}{\mathbf{T^2}_{l,\alpha}} \tag{2.30}$$

Theoretical thresholds have likewise been proposed for these statistics, see [25].

2.5.2.2 Angular-based statistics

The second main approach to fault detection using principal components variables consists in monitoring the direction of the principal subspace through angular-based indices [51, 52]. A fault, particularly a structural one, in a system modifies the process dynamic behaviour which is characterised by the principal subspace. So if we consider that the principal subspace estimated from the reference fault-free observation set is denoted $\hat{S}_r$ and the one estimated from the current observation set is denoted $\hat{S}_c$, a structural fault may be detected by comparing the direction of $\hat{S}_c$ against $\hat{S}_r$. A monitoring index based on the inner product of loading vectors, i.e. the angle Ω between principal component directions can be used for this purpose, see Fig.2.3. The principal components in such case may be updated recursively [53, 54] for example.

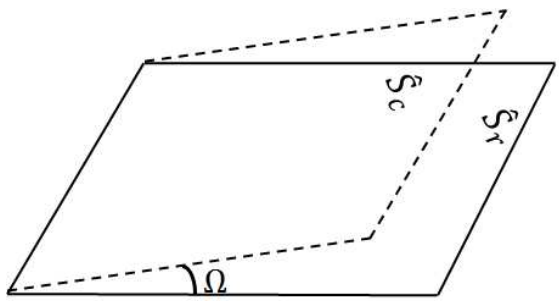

Figure 2.3: Angular-based fault detection

Even if this approach is successful when the distance-based statistics fail, a fault will not be detected by the angular index, unless it makes a meaningful change in the correlation structure so that it captures one or more principal directions. In [51], this methodology is illustrated on multivariate autoregressive (AR) models where the fault is considered as a mean step change in one monitored variable, or a parameter change in the model. The range of the fault is always above 10%. It has been shown that detecting small shifts requires monitoring directions of the last principal components rather than the first ones. The superiority of dynamic monitoring with this approach over the static monitoring was proved. However, the detection results of small changes were not satisfactory. Moreover, the detection performance depends strongly on the choice of the sample-window size and the number of retained principal components.

2.5.3 Application of KL divergence to the principal components

The main shortcoming of the statistical indices used for fault detection with PCA, is that they treat each observation individually and don't include information from past data. That makes them insensitive to small shifts in the process variables especially for small fault-to-noise ratio (close to 0 dB), and causes many false alarms. This is in general the disadvantage of Schewhart control charts, including $\mathbf{T^2}$ and SPE. EWMA, CUSUM and GLR-based control charts address this problem by accumulating information from acquired last observations [55, 56, 57]. They are

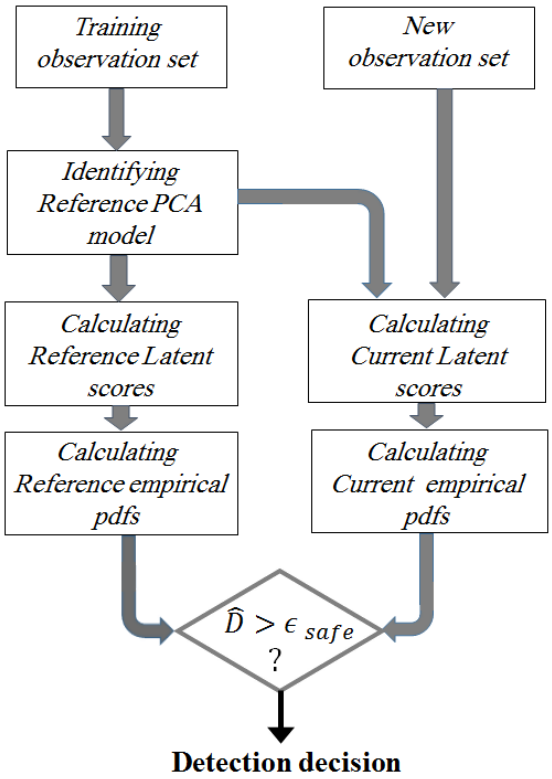

Figure 2.4: Divergence-based fault detection

commonly used in model-based FDD framework, where they operate on generated residuals to detect a fault change with a known profile. Here we propose to detect slight changes caused by faults, using KL divergence estimated through Eq.2.9 and applied to the latent principal component scores. The probability distribution of the first principal components will be monitored, without worrying about the change profile to be detected and the probability distributions form.

The block diagram in Fig.2.4 shows the main steps to make decision about the

fault occurrence. The PCA model should be identified from a training observation set obtained when the process is operating under fault-free conditions. The principal subspace extracted with classical PCA assumes that the first two statistical moments of data are time-independent. This means that the process generating the data is supposed stationary. Therefore, for non-stationary systems with several operating points, the principal subspace will be associated with only one operating point. One can imagine, however, that a bank of principal subspaces may be constructed for a set of operating points. The fault detection, then, uses the reference principal subspace associated with the current operating point. Afterwards, the latent principal scores are calculated by projecting the fault-free observations into the principal subspace, and their empirical pdfs are estimated. For each new observation set, the latent scores are calculated using the PCA model, and the new pdfs are estimated. $\hat{D}$ is calculated between the reference and the current pdfs according to Eq.2.9. Decision is made by comparing $\hat{D}$ to an appropriate threshold.

2.6 Evaluation on a numerical example

2.6.1 Fault detection procedure description

The KL divergence has been applied to monitor a multivariate process from which are collected measurements of 8 correlated variables inspired from [58] and defined at instant i as follows :

$$\begin{cases} x_1(i) = 1 + \sin(0.1i) \\ x_2(i) = 2\cos(i/4)^3 \exp(-i/N) \\ x_3(i) = \log(x_2(i)^2) \\ x_4(i) = x_1(i) + x_2(i) \\ x_5(i) = x_1(i) - x_2(i) \\ x_6(i) = 2x_1(i) + x_2(i) \\ x_7(i) = x_1(i) + x_3(i) \\ x_8(i) \propto \mathcal{N}(0,1) \end{cases}$$

The equations above show linear and non-linear relationships between variables, and also an independent variable. N observations are generated for each variable, to which a zero mean random noise drawn from normal distribution is added to simulate measurement noises and errors. Each variable is autoscaled to have zero mean and unit variance, and the $\bar{X}$ matrix is formed. The eigenvector decomposition of the data correlation matrix gives a principal subspace spanned by four principal components. The principal subspace dimension can be directly identified from the equations above showing 4 linear relationships that lead to a 4-dimension residual subspace.

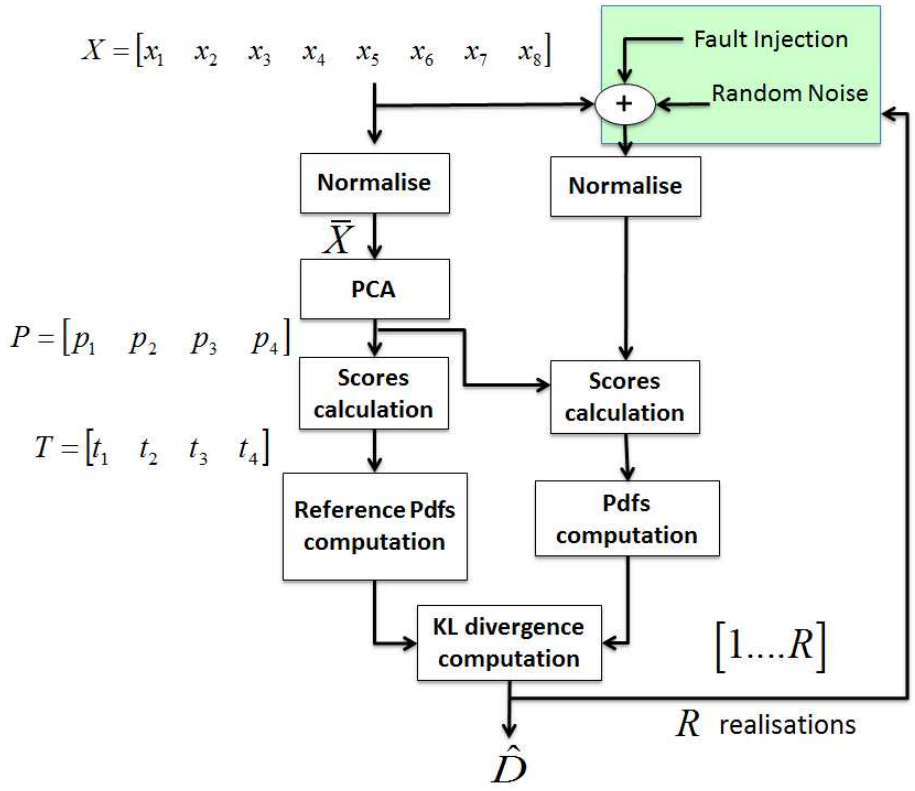

Figure 2.5: Procedure description

To test the divergence capability for detecting faults upon this process, the following procedure which is illustrated in Fig.2.5 is adopted:

1. The PCA model is established and 4 principal components are retained to span the principal subspace.

2. The scores matrix is obtained from the data matrix $\bar{X}$ and the loading matrix P using $T = \bar{X}P$. The first 4 columns of T are the latent scores while the remaining ones are the residual.

3. The probability density of each of the 4 latent scores is estimated using a normal kernel estimator. The obtained pdfs are set as references of the normal operating conditions of the process.

4. Another $\bar{X}$ matrix is generated with a different added noise. The corresponding latent scores and their pdfs are calculated.

5. The divergence is then calculated between these pdfs and their references. Steps 4-5 are repeated for several generated matrices, while introducing a fault into some matrices.

We will analyse the detection results for different SNR, data sizes N and fault amplitudes, and then we will illustrate the effect of different SNR and N values on the fault detectability. A general fault model is considered: the change due to the fault is assumed as a bias that depends on the signal amplitude. The ith observation from the variable $\boldsymbol{x_j}$ is written as:

$$x_{ij} = x^*_{ij} + a \times x^*_{ij} \tag{2.31}$$

where the star (*) mark denotes the fault-free and noiseless observations. a is the fault amplitude: $a \neq 0$ at instant i if $\boldsymbol{x_j}$ is faulty. As the fault is considered incipient, only the few last observations generated are considered faulty, see Fig.2.6. In the following, detection results are shown for a proportion of faulty observations $(N - b)/N$ equal to 10%. Then different proportions will be considered in the evaluation of the divergence detection performance. As we are particularly interested

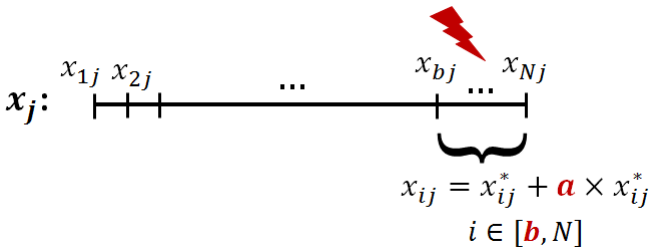

Figure 2.6: Fault description

in incipient faults that cause small changes with respect to the signal variations, the range [0,10%] for the fault parameters a and $(N-b)/N$ is our concern. The generated observations are corrupted with an additive white Gaussian noise (AWGN). The noisy observation of the faulty variable $\boldsymbol{x_j}$ at instant i is $x_{ij} = x^*_{ij} + a \times x^*_{ij} + v_{ij}$, where v_{ij} is the noise sample drawn from the distribution $\mathcal{N}(0, \sigma_v^2)$. For a specific SNR, the FNR can be calculated from Eq.2.14, knowing that $\sigma_f^2 = \dfrac{a}{N}\sum_{i=b}^{N}(x^*_{ij} - \mu_j)^2$ where μ_j is the sample mean of $\boldsymbol{x_j}$.

2.6.2 Detection results

2.6.2.1 Detection with common statistics

Let's consider that $N = 1000$ samples are generated for each variable and corrupted by a random noise that provides a SNR= 35 dB. Fig.2.7 shows the different variables. Applying PCA to these variables yields 8 new variables, the principal

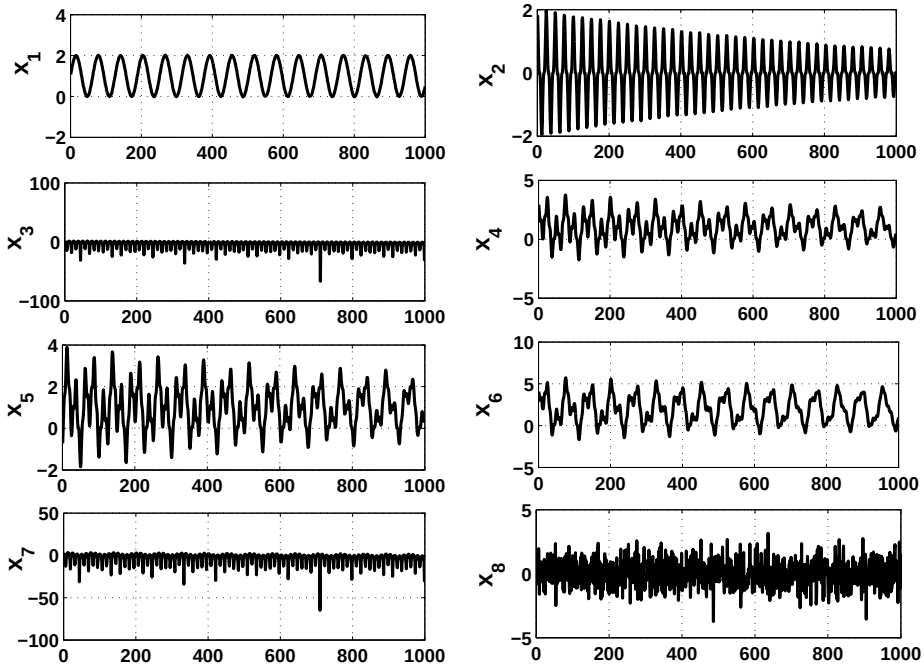

Figure 2.7: Process variables

component scores, shown in Fig.2.8. The first scores $\boldsymbol{t_1}$, $\boldsymbol{t_2}$, $\boldsymbol{t_3}$ and $\boldsymbol{t_4}$ constitute the principal subspace, which represent 99% of data variations. To test the Hotelling $\mathbf{T^2}$ capability for detecting faults in this process, we corrupt the last 100 observations of the variable $\boldsymbol{x_1}$ by a large bias fault equal to 150% of observation amplitude ($a = 1.5$). The $\mathbf{T^2}$ test does not provide any indication of the fault presence, as shown in Fig.2.9. The detection is made with respect to two different thresholds. The horizontal dashed line represents the practical threshold computed as the maximal 99% of the nominal $\mathbf{T^2}$, and the other horizontal line is the theoretical threshold calculated at significance level 0.05. The latter is regularly crossed inducing a high false alarm rate. Besides, most of the $\mathbf{T^2}$ values in the sampling interval [900:1000] delimited by the dashed rectangle are relatively weak, that makes the fault detection unreliable. This result explains the general tendency in the literature to use the residual subspace rather than the principal one in order to detect faults. The $\mathbf{T^2}$

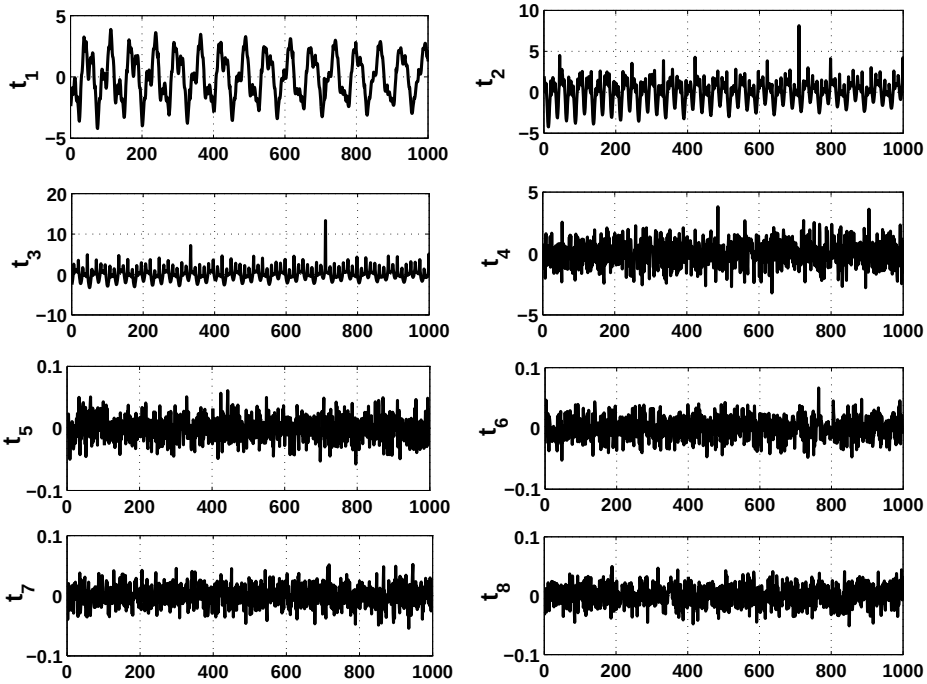

Figure 2.8: Principal component scores

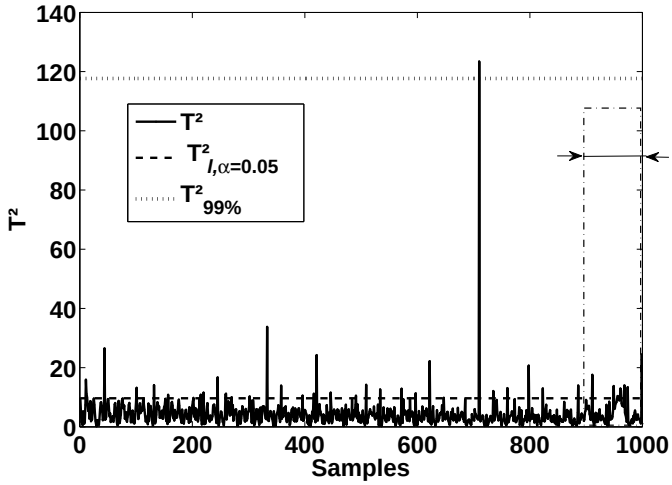

Figure 2.9: Fault detection result of a large bias (150%) on $\boldsymbol{x_1}$ using $\mathbf{T^2}$ statistic

fails to detect the faults because of the large amount of variability naturally present in the principal subspace [59].

Fig.2.10 depicts the behaviour of SPE when a 10% bias of amplitude, i.e. $a = 0.1$, affects the last 100 observations of $\boldsymbol{x_1}$. The horizontal line represents

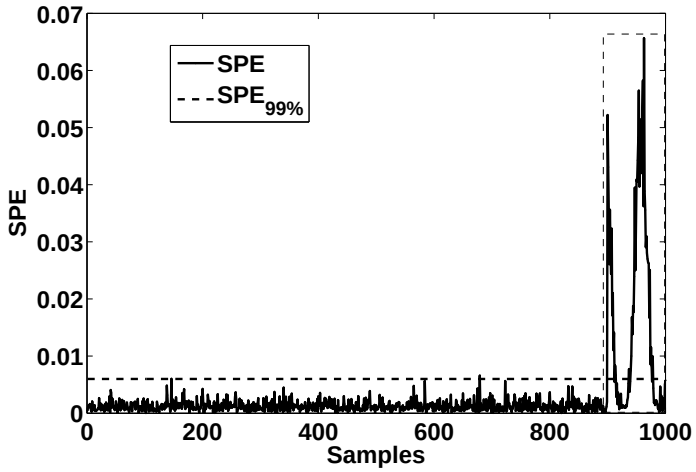

Figure 2.10: Fault detection result of a bias of 10% on $\boldsymbol{x_1}$ using SPE statistic

the maximal 99% of the nominal SPE value. The fault is readily detectable by the SPE, even though some observations in the faulty interval do not exceed the SPE threshold. This is due to the fact that the SPE is very sensitive to the observation magnitude. The fault should be serious enough so that the SPE increases value and crosses the given threshold. Moreover, we show in Fig.2.11 the detection of a

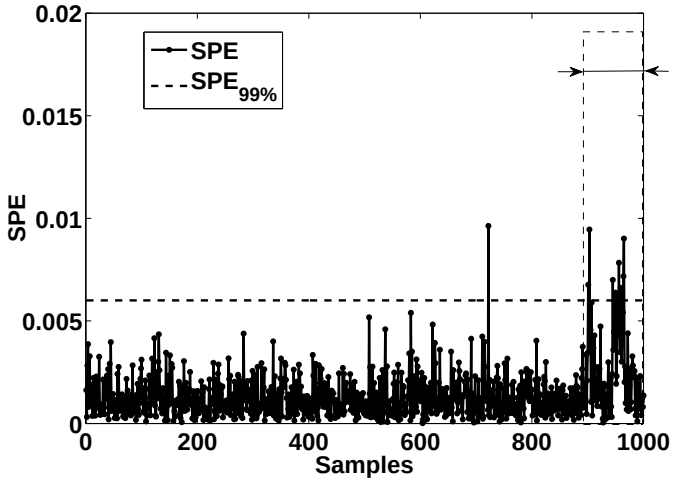

Figure 2.11: Fault detection result of a bias of 3% on $\boldsymbol{x_1}$ using SPE statistic

3% bias that affects x_1 ($a = 0.03$). Only 9 out of 100 faulty observations exceed the

SPE threshold, while a lower threshold will cause numerous false alarms.

As shown by these results, the common statistics ($\mathbf{T^2}$ and SPE) exhibit low capability for incipient fault detection. In fact, using these features leads to poor performance in terms of false alarms and missed detections. In the next paragraph, we will evaluate the KL divergence capability.

2.6.2.2 Detection with KL divergence

Using the Kullback-Leibler divergence, we expect to enhance the fault detection efficiency in the principal subspace. It allows evaluating the distributional dissimilarity of principal scores before and after the fault. It assumes that the two distributions share the same support, and the probability distributions are stationary, i.e. time-independent. The single fault case is considered to show the fault detection capability of the divergence for each variable. The last 100 out of $N = 1000$ observations generated from variables are considered faulty. The SNR is still set to 35 dB.

Fig.2.12 displays the probability density of the latent score $\boldsymbol{t_1}$ before and after a large bias fault (150%) affecting $\boldsymbol{x_1}$. One can notice the difference between the two pdfs. The difference will be slighter for a smaller fault, but the divergence will emphasize it. To illustrate this fact, 100 realizations of the data matrix X are made, among them we introduce a fault bias of 10% into one of the variables. The

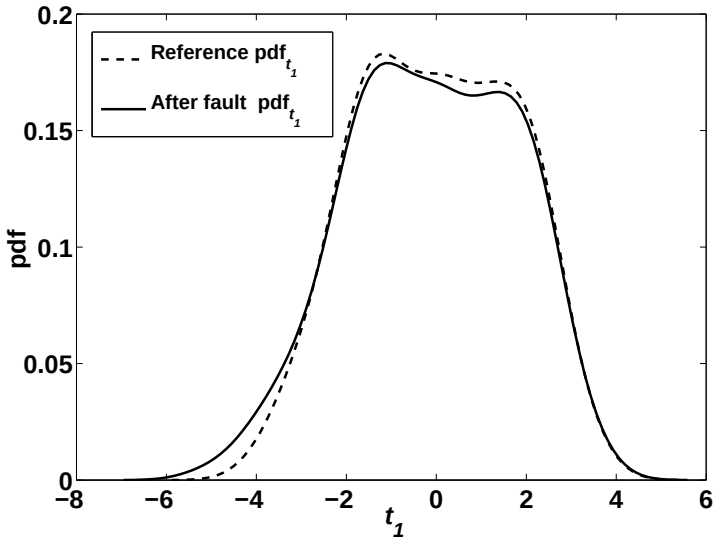

Figure 2.12: Probability density of $\boldsymbol{t_1}$ before and after a 150% bias fault on $\boldsymbol{x_1}$

divergence is calculated for each of the latent scores according to Eq.2.9. A fault on $\boldsymbol{x_1}$ has been introduced into some realisations of the data matrix, precisely into

the 20th-25th realisations, the 60th, 75th and the 95th-100th realisations. Fig.2.13

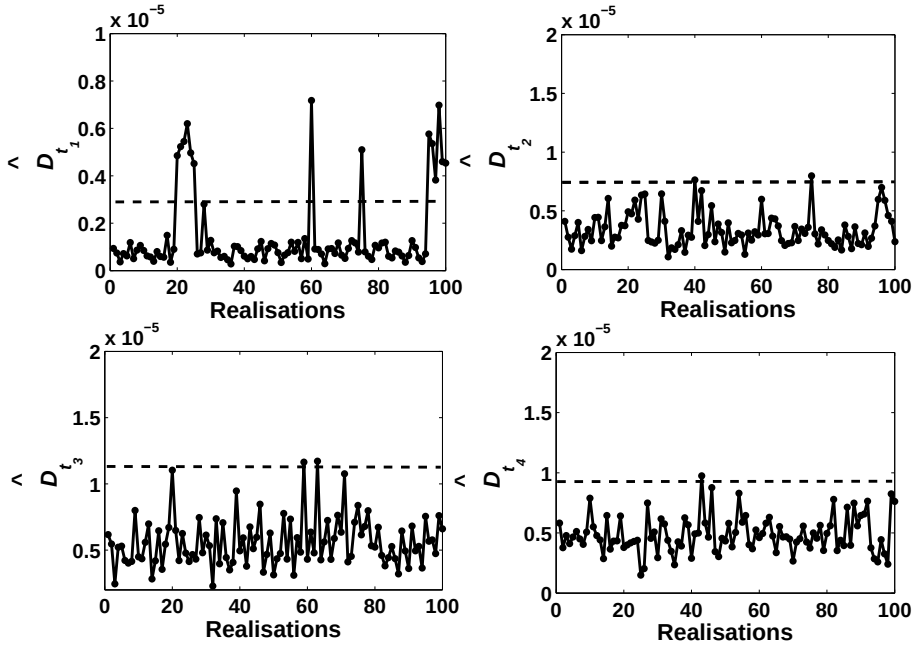

Figure 2.13: Fault detection result of a 10% bias fault on $\boldsymbol{x_1}$ using KL divergence

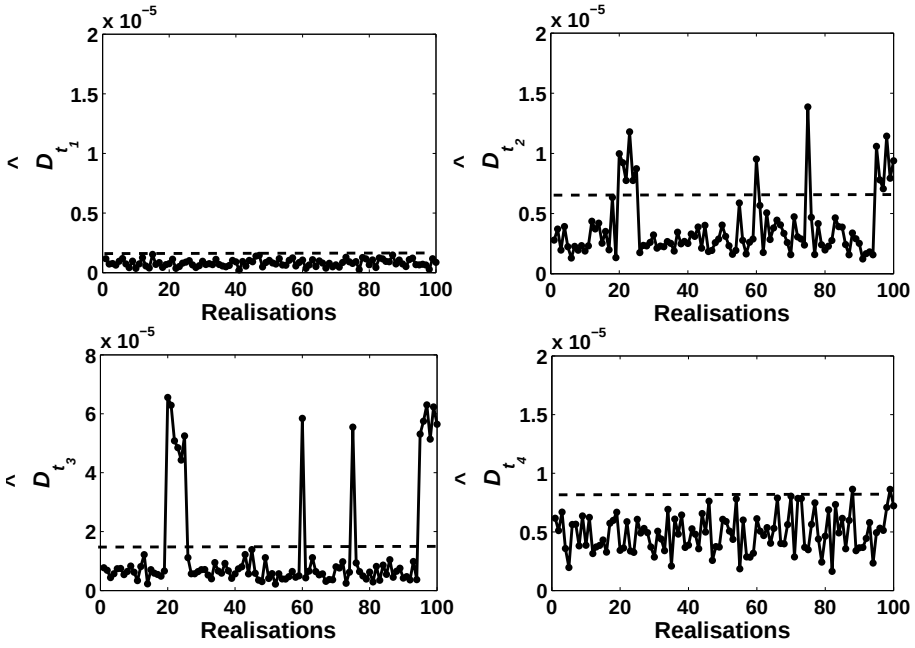

Figure 2.14: Fault detection result of a 10% bias fault on $\boldsymbol{x_3}$ using KL divergence

shows that the fault that affects $\boldsymbol{x_1}$ is detected by the divergence computed for $\boldsymbol{t_1}$, while $\boldsymbol{t_2}$, $\boldsymbol{t_3}$ and $\boldsymbol{t_4}$ are still not sensitive. The first plot shows 14 high $\hat{D}$ values, which cross the corresponding threshold. This complies with the expected result, because these points correspond to faulty realisations. If the variable $\boldsymbol{x_3}$ is faulty, the fault impact would be shown on the $\hat{D}$ value computed for $\boldsymbol{t_2}$ and $\boldsymbol{t_3}$, as displayed in Fig.2.14.

Similar results, summarised in Table 2.1, are obtained for the faults on the other variables. Grey cells indicate the successful detections. We show for example that the divergence computed for $\boldsymbol{t_1}$ detects the faults that affect $\boldsymbol{x_1}$, $\boldsymbol{x_4}$, and $\boldsymbol{x_6}$. Faults on all variables can be detected by the divergence computed on the latent scores.

Table 2.1: Faults discrimination capability of the KL divergence [60]

	δx_1	δx_2	δx_3	δx_4	δx_5	δx_6	δx_7	δx_8
$\boldsymbol{t_1}$	■			■		■		
$\boldsymbol{t_2}$		■	■		■		■	
$\boldsymbol{t_3}$			■		■		■	
$\boldsymbol{t_4}$								■

The KL divergence value $\hat{D}$ is also able to characterise the fault severity: the

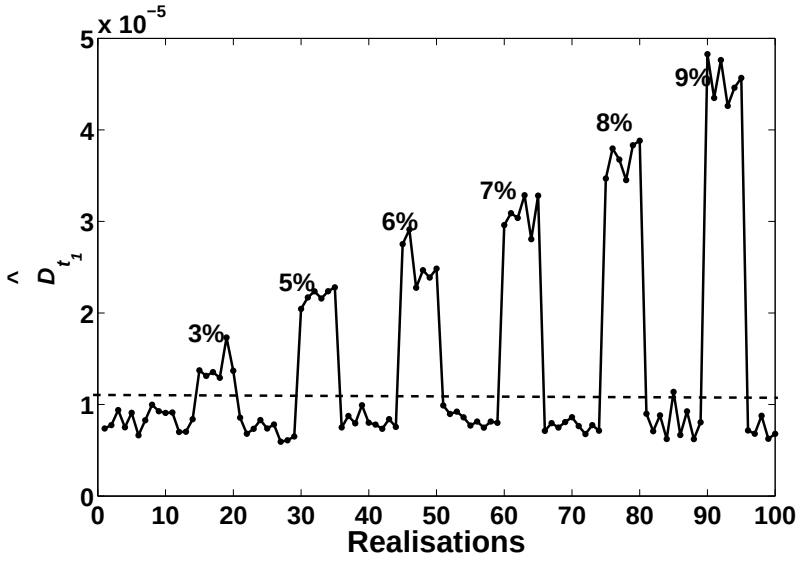

Figure 2.15: Divergence sensitivity to the fault amplitude

higher the fault level, the higher the disparity between the current pdf and the reference one. As we are interested in detecting small faults, we introduce into some

data matrices different low bias that affect the last observations of $\boldsymbol{x_1}$: a bias of 3% into the 15th-20th matrices/realisations, 5% for the 30th-35th, 6% for the 45th-50th, 7% for the 60th-65th matrices, 8% for the 75th-80th and 9% for the 90th-95th. The result is depicted in Fig.2.15. It clearly shows that $\hat{D}$ is an effective image of the fault severity. $\hat{D}$ is sensitive to the fault level, and therefore, one may refer to the divergence computed for the latent scores to follow the evolution of an abnormal event and make appropriate decisions.

2.6.3 Detection performance evaluation with respect to small faults

2.6.3.1 The evaluation procedure

The fault described in Fig.2.6 is used in the evaluation. As we are particularly interested in incipient faults that cause small changes with respect to the signal variations, the range [0,10%] for the fault parameters a and $(N-b)/N$ is our concern. The FNR can hence be calculated from Eq.2.14, knowing that $\sigma_f^2 = \frac{a}{N}\sum_{i=b}^{N}(x_{ij}^* - \mu_j)^2$ where μ_j is the sample mean of $\boldsymbol{x_j}$. Fig.2.16 displays the region of FNR associated

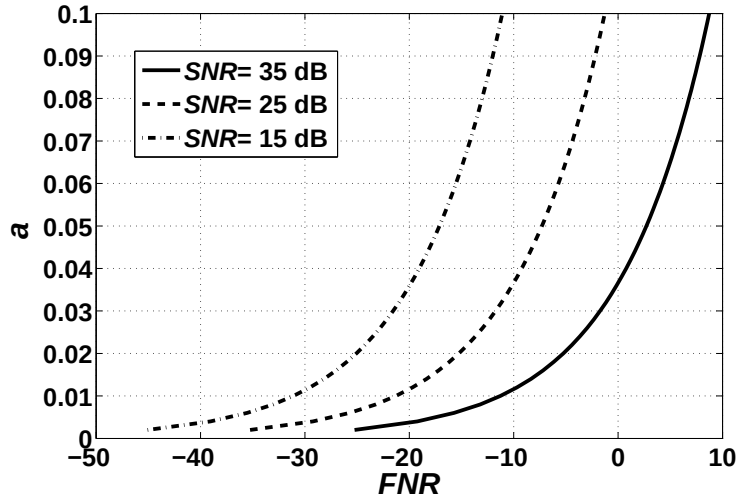

Figure 2.16: FNR range associated to small faults

with such faults, here $j = 1$, $a \in [0, 10\%]$, $(N-b)/N = 0.1$ and $N = 1000$. It is shown that for such faults, the FNR lies in the interval $]-\infty, 10]$ dB.

We intend in this part of the work to evaluate the detection error probabilities to prove the efficiency of the proposed technique. In order to evaluate the

detection error rates, $\hat{D}$ has been shown to approximately fit a Gamma distribution with shape and scale parameters that are specific to the noise and fault conditions. Fig.2.17 shows the histogram of $\hat{D}$ in fault-free case for $SNR = 35$ dB. The Statistics/Probability Distributions toolbox of Matlab is used to compute the threshold required for a specific false alarm rate P_{FA}, by first estimating Gamma distribution parameters from divergence data, and then obtaining the gamma inverse cumulative distribution at the $1 - P_{FA}$ value.

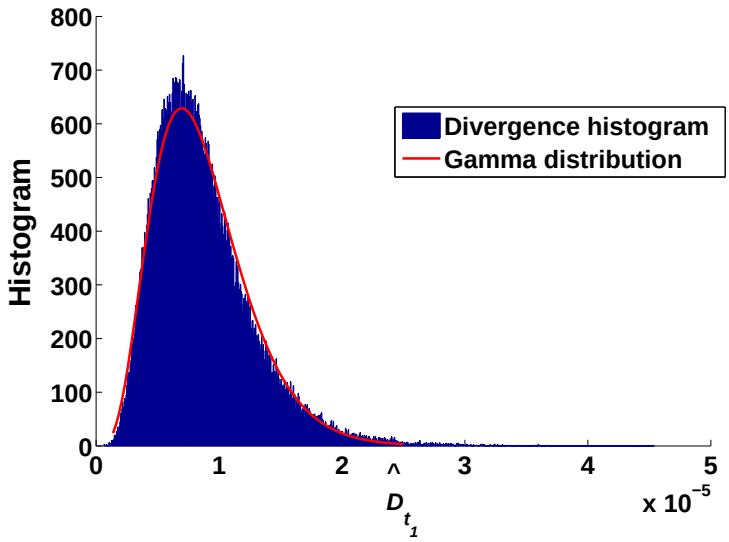

Figure 2.17: Gamma distribution fitting to divergence histogram

Usually, the threshold for detection is chosen in such a way to achieve a balance between false alarm and missed detection rates. As we aim here to evaluate the fault detection capabilities of the divergence, we will evaluate the missed detection rate that can be achieved given a low or trivial false alarm rate. Results will be shown for the variable $\boldsymbol{x_1}$ affected by the fault, in function of the fault amplitude as well as the FNR, for different SNR and different sample sizes N.

2.6.3.2 Evaluation results

P_{FA} is set to 0.05 which is a low and practical rate of false alarm. The fault is characterised with two parameters, namely the proportion of the faulty sample $(N-b)/N$ and the fault amplitude a. The detection performance will therefore depend upon these two parameters. Fig.2.18 depicts, for SNR= 35 dB and N= 1000, the P_{MD} evaluated with respect to a for three sizes of the faulty sample. In the most tedious case, where $(N-b)/N$= 0.05, the divergence is able to detect a fault which amplitude is equal to 6% of the signal magnitude with negligible error probabilities. The

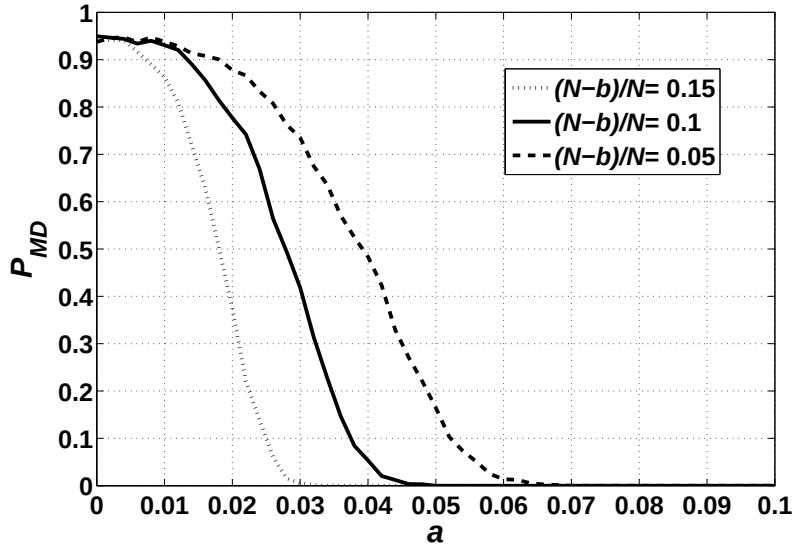

Figure 2.18: P_{MD} versus the fault amplitude a, SNR= 35 dB, N= 1000

sensitivity to smaller a (3%) rises when $(N-b)/N$ increases because the latter reflects the fault impact. Fig.2.19 illustrates the same P_{MD} variation but with respect

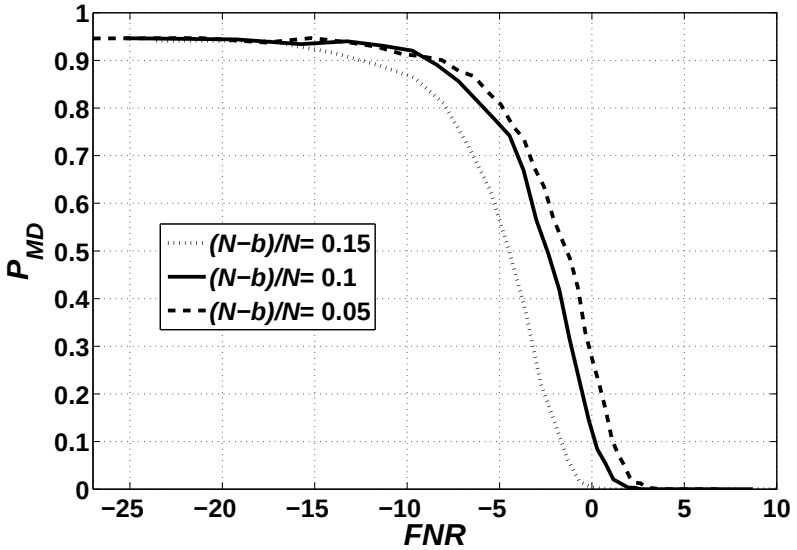

Figure 2.19: P_{MD} versus FNR, SNR= 35 dB, N= 1000

to FNR. It is shown that the null P_{MD} is obtained for FNR around 0 dB.

A lower SNR will certainly reduce the sensitivity of the divergence to small fault

amplitude. However, a high N can notably mitigate the impact of a high noise level. We illustrate in Fig.2.20 the P_{MD} obtained with SNR= 25 dB and $(N-b)/N$= 0.1 while considering different N. The result is likewise depicted, in Fig.2.21, in func-

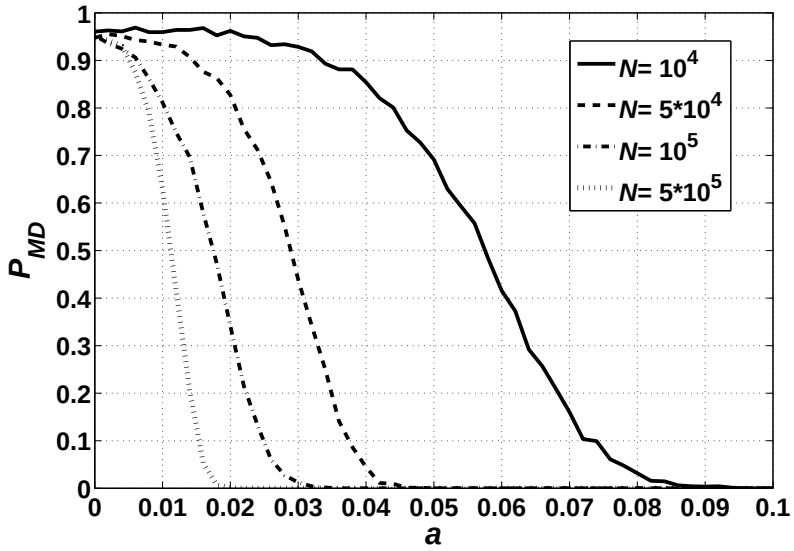

Figure 2.20: P_{MD} versus the fault amplitude a, SNR= 25 dB, $(N-b)/N$=0.1

tion of FNR. It is shown that very small a (2%) to which is associated a negative FNR (-15 dB) can be detected (P_{MD}=0.001) in this case while N is sufficiently large, equal to $5*10^5$. For lower N, 10^4 observations, the divergence is still able to detect 9% change of the signal magnitude with zero missed detection probability.

On the other hand, Fig.2.22 displays the missed detection rate obtained with the SPE criterion for the same prefixed false alarm rate ($P_{FA}=0.05$). The result shown in the plot is independent from N and this is normal because the SPE treats each observation individually. The SPE detects small faults whose amplitude $a<0.1$ with a $P_{MD}>0.3$. The divergence has shown better performance. Obviously, the detection of small amplitude shifts with the divergence is effective as much as the noise and data conditions are favorable (low noise level, high SNR, large N). The impact of a low SNR on the divergence detection performance can however be mitigated by using a large data size, thanks to its global character. Incipient faults with negative FNR can so be successfully detected with low detection error probabilities.

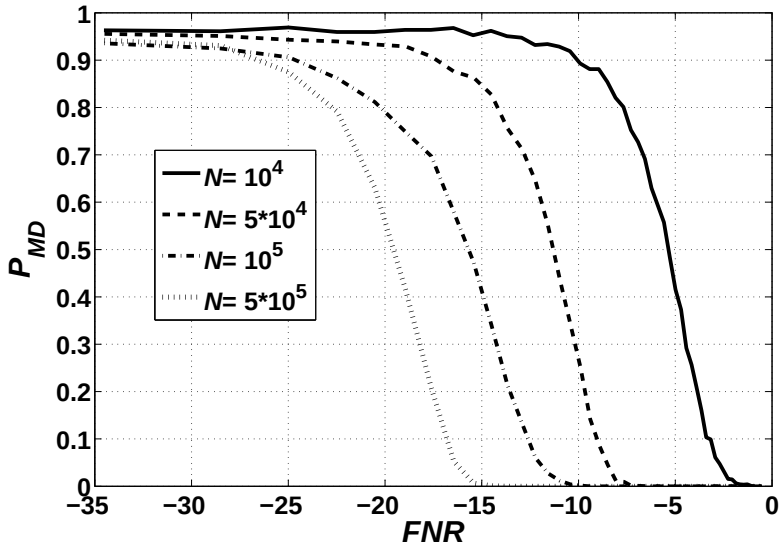

Figure 2.21: P_{MD} versus FNR, SNR= 25 dB, $(N-b)/N$=0.1

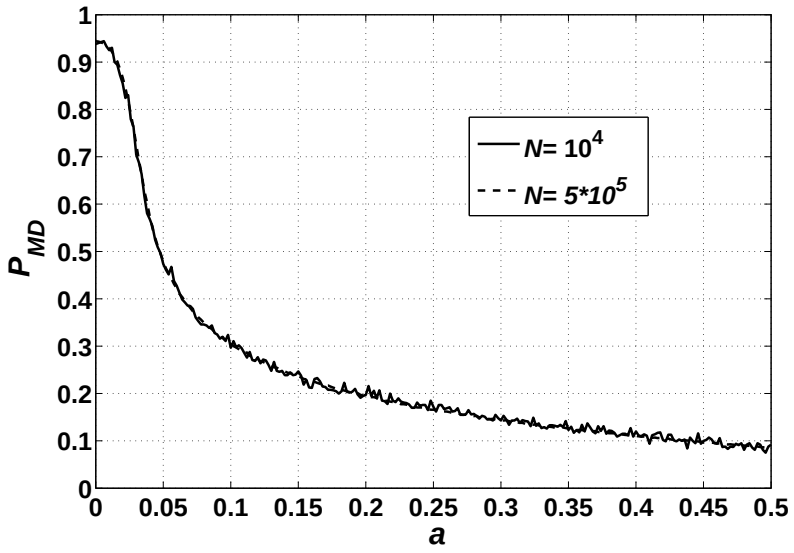

Figure 2.22: P_{MD} obtained with the SPE, SNR= 25 dB

2.7 Conclusion

The evaluation of the common statistics used in the PCA framework ($\mathbf{T^2}$ and SPE) has shown poor detection capability in the presence of incipient faults. Therefore, the KL divergence has been proposed to be a general distribution-free fault indicator, characterised by high sensitivity with respect to incipient faults (the short

duration change (10% of sample size) whose amplitude is less than 10% of the signal magnitude). In the presence of high-level noise ($SNR \leq 25$ dB), the divergence needs to be computed on a sample size large enough ($N > 10^4$) so that high fault detection performance ($P_{FA} = 5\%$ and $P_{MD} = 0.1\%$) can be achieved. Beside fault detection capability, the divergence has also shown its capability to reflect the fault severity (amplitude).

The next chapter is concerned with proving this statement through a theoretical study that aims at developing an analytical model that explicitly expresses the divergence in function of the fault parameters.

Bibliography

[1] C. Svard, M. Nyberg, E. Frisk, and M. Krysander, "Data-driven and adaptive statistical residual evaluation for fault detection with an automotive application," *Mechanical Systems and Signal Processing*, vol. 45, no. 3, pp. 170–192, 2014.

[2] W. Zhou, T. Habetler, and R. Harley, "Bearing fault detection via stator current noise cancellation and statistical control," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4260–4269, 2008.

[3] D. Garcia-Alvarez, M. J. Fuente, and G. Sainz, "Design of residuals in a model-based fault detection and isolation system using statistical process control techniques," *IEEE 16th Conference on Emerging Technologies & Factory Automation (ETFA)*, pp. 1–7, 2011.

[4] C.-M. Fan, R.-S. Guo, S.-C. Chang, and C.-S. Wei, "Shewma: an end-of-line SPC scheme using wafer acceptance test data," *IEEE Transactions on Semiconductor Manufacturing*, vol. 13, no. 3, pp. 344–358, 2000.

[5] X. Shibo, L. Yuansheng, and L. Mei, "Quality control charts for log-normal distribution based on bootstrap method," *27th Chinese Control Conference*, pp. 18–21, 2008.

[6] T. Lazariv, W. Schmid, and S. Zabolotska, "On control charts for monitoring the variance of a time series," *Journal of Statistical Planning and Inference*, vol. 143, no. 9, pp. 1512–1526, 2013.

[7] M. Yun and Z. Youlin, "Q control charts for negative binomial distribution," *Computers & Industrial Engineering*, vol. 31, no. 3-4, pp. 813–816, 1996.

[8] M. A. Graham, A. Mukherjee, and S. Chakraborti, "Distribution-free exponentially weighted moving average control charts for monitoring unknown location," *Computational Statistics & Data Analysis*, vol. 56, no. 8, pp. 2539–2561, 2012.

[9] N. Balakrishnan, I. S. Triantafyllou, and M. V. Koutras, "Nonparametric control charts based on runs and wilcoxon-type rank-sum statistics," *Journal of Statistical Planning and Inference*, vol. 139, no. 9, pp. 3177–3192, 2009.

[10] S. T. Bakir, "A distribution-free shewhart quality control chart based on signed-ranks," *Quality Engineering*, vol. 16, no. 4, pp. 613–623, 2004.

[11] ——, "A nonparametric shewhart-type quality control chart for monitoring broad changes in a process distribution," *International Journal of Quality, Statistics, and Reliability*, vol. 2012, pp. 1–10, 2012.

[12] R. J. Patton, P. M. Frank, and R. N. Clark, *Issues of Fault Diagnosis for Dynamic Systems*. Berlin Heidelberg: Springer, 2000.

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

[14] D. Romano and M. Kinnaert, "Robust fault detection and isolation based on the kullback divergence," *Fault Detection, Supervision and Safety of Technical Processes*, vol. 1, pp. 426–431, 2006.

[15] C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon, "A multivariate exponentially weighted moving average control chart," *Technometrics*, vol. 34, pp. 46–53, 1992.

[16] S. W. Cheng and K. Thaga, "Single variables control charts : an overview," *Quality and Reliability Engineering International*, vol. 22, pp. 811–820, 2006.

[17] D. M. Hawkins and E. M. Maboudou-Tchao, "Multivariate exponentially weighted moving covariance matrix," *Technometrics*, vol. 50, pp. 155–166, 2008.

[18] J. Tajer, A. Makke, O. Salem, and A. Mehaoua, "A comparison between divergence measures for network anomaly detection," *7th International Conference on Network and Service Management (CNSM)*, pp. 1–5, 2011.

[19] O. Salem, F. Nait-Abdesselam, and A. Mehaoua, "Anomaly detection in network traffic using jensen-shannon divergence," *IEEE International Conference on Communications (ICC)*, pp. 5200–5204, 2012.

[20] Y. Qiao and N. Minematsu, "A study on invariance of f -divergence and its application to speech recognition," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3884–3890, 2010.

[21] W. L. Hung and M. S. Yang, "On the j-divergence of intuitionistic fuzzy sets with its application to pattern recognition," *Information Sciences*, vol. 178, pp. 1641–1650, 2008.

[22] M. Basseville, "Divergence measures for statistical data processing : An annotated bibliography," *Signal Processing*, vol. 93, pp. 621–633, 2013.

[23] E. Mwebaze, P. Schneider, F. M. Schleif, J. R. Aduwo, J. A. Quinn, S. Haase, T. Villmann, and M. Biehl, "Divergence-based classification in learning vector quantization," *Neurocomputing*, vol. 74, pp. 1429–1435, 2011.

[24] M. Basseville, "Detecting changes in signals and systems : A survey," *Automatica*, vol. 24, pp. 309–326, 1988.

[25] Y. Tharrault, "Diagnostic de fonctionnement par analyse en composantes principales: application à une station de traitement des eaux usées," Ph.D. dissertation, National Polytechnic Institute of Lorraine, 2008.

[26] C. Sankavaram, B. Pattipati, K. Pattipati, Y. Zhang, M. Howell, and M. Salman, "Data-driven fault diagnosis in a hybrid electric vehicle regenerative braking system," *IEEE Aerospace Conference*, pp. 1–11, 2012.

[27] Y. Gao, T. Yang, N. Xing, and M. Xu, "Fault detection and diagnosis for spacecraft using principal component analysis and support vector machines," *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1984–1988, 2012.

[28] M. Hamadache and D. Lee, "Principal components analysis based fault detection and isolation for electronic throttle control system," *12th International Conference on Control, Automation and Systems (ICCAS),*, pp. 808–813, 2012.

[29] A. A. Silva, A. M. Bazzi, and S. Gupta, "Fault diagnosis in electric drives using machine learning approaches," *IEEE International Electric Machines & Drives Conference (IEMDC)*, pp. 722–726, 2013.

[30] J. Yu, "Fault detection using principal components-based gaussian mixture model for semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, no. 3, pp. 471–486, 2011.

[31] F. M. Coetzee, "Correcting the Kullback-Leibler distance for feature selection," *Pattern Recognition Letters*, vol. 26, pp. 1675–1683, 2013.

[32] W. Zhang, S. Shan, X. Chen, and W. Gao, "Local gabor binary patterns based on Kullback-Leibler Divergence for partially occluded face recognition," *IEEE Signal Processing Letters*, vol. 14, pp. 875–878, 2007.

[33] A. Anderson and H. Haas, "Kullback-Leibler Divergence (KLD) based anomaly detection and monotonic sequence analysis," *IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, 2011.

[34] F. Barbir, "Chapter eight - fuel cell diagnostics," in *PEM Fuel Cells, Theory and Practice (Second Edition)*, 2013, pp. 265–304.

[35] Y. Xiang, K. Li, and W. Zhou, "Low-rate ddos attacks detection and traceback by using new information metrics," *IEEE Transactions on Information Forensics and Security*, vol. 6, pp. 426–437, 2011.

[36] J. Silva and S. Narayanan, "Average divergence distance as a statistical discrimination measure for hidden Markov models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 890–906, 2006.

[37] A. Gupta, S. Parameswaran, and C.-H. Lee, "Classification of electroencephalography (EEG) signals for different mental activities using Kullback-Leibler (KL) divergence," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1697–1700, 2009.

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

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

[40] ——, "Chapter three - properties of sample principal components," in *Principal Component Analysis (Second Edition)*, 2002, pp. 265–304.

[41] M. Tamura and S. Tsujita, "A study on the number of principal components and sensitivity of fault detection using PCA," *Computers & Chemical Engineering*, vol. 31, pp. 1035–1046, 2007.

[42] M. Kano, K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss, and B. R. Bakshi, "Comparison of multivariate statistical process control monitoring methods with applications to the eastman challenge problem," *Computers & Chemical Engineering*, vol. 26, pp. 161–174, 2002.

[43] E. R. Malinowski, *Factor Analysis in Chemistry*. Wiley-Interscience, New York, 1991.

[44] S. Wold, "Cross-validatory estimation of the number of components in factor and principal components models," *Technometrics*, vol. 20, 1978.

[45] S. J. Qin and R. Dunia, "Determining the number of principal components for best reconstruction," *Journal of Process Control*, vol. 10, pp. 245–250, 2000.

[46] S. Valle, W. Li, and S. J. Qin, "Comparison of multivariate statistical process control monitoring methods with applications to the eastman challenge problem," *Industrial & Engineering Chemistry Research*, vol. 38, pp. 4389–4401, 1999.

[47] B. Mnassri, E. M. E. Adel, and M. Ouladsine, "Generalization and analysis of sufficient conditions for pca-based fault detectability and isolability," *Annual Reviews in Control*, vol. 37, pp. 154–162, 2013.

[48] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems : Effect of inequality of variance in one-way classification," *The Annals of Mathematical Statistics*, vol. 25, pp. 290–302, 1954.

[49] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recognition Letters*, vol. 25, pp. 711–724, 2004.

[50] H. Yue and S. Qin, "Reconstruction-based fault identification using a combined index," *Industrial & Engineering Chemistry Research*, vol. 40, pp. 4403–4414, 2001.

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

[52] A. Raich and A. Cinar, "Diagnosis of process disturbances by statistical distance and angle measures," *Computers & Chemical Engineering*, vol. 21, pp. 661–673, 1997.

[53] W. Li, H. Yue, S. Valle-Cervantes, and S. Qin, "Recursive pca for adaptive process monitoring," *Journal of Process Control*, vol. 10, pp. 471–486, 2000.

[54] S. C. Chan, H. C. Wu, and K. M. Tsui, "Robust recursive eigendecomposition and subspace-based algorithms with application to fault detection in wireless sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 1703–1718, 2012.

[55] M. Gálvez-Carrillo and M. Kinnaert, "Sensor fault detection and isolation in doubly-fed induction generators accounting for parameter variations," *Renewable Energy*, vol. 36, pp. 1447–1457, 2011.

[56] F. Gustafsson, "Statistical signal processing approaches to fault detection," *Annual Reviews in Control*, vol. 31, no. 1, pp. 41–54, 2007.

[57] S. Chitraganti, S. Aberkane, and C. Aubrun, "Statistical properties of exponentially weighted moving average algorithm for change detection," *51st Annual Conference on Decision and Control (CDC)*, pp. 574–578, 2012.

[58] 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.

[59] J. Harmouche, C. Delpha, and D. Diallo, "Faults diagnosis and detection using principal component analysis and Kullback-Leibler divergence," *8th Annual Conference on IEEE Industrial Electronics Society (IECON)*, 2012.

[60] ——, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part I," *Elsevier Signal Processing*, vol. 94, pp. 278–287, 2014.

Chapter 3

Kullback-Leibler Divergence for Fault Estimation

3.1 Introduction

The fault estimation problem has gained considerable attention in recent years. If a fault in the process measurements has been detected, and the information contained into the data is important, it is necessary to retrieve the fault-free measurements from the faulty ones [1, 2, 3]. Sensor validation and correction is concerned with the problem of identifying the fault magnitude in order to retrieve the sensor response from faulty sensor data [4, 5]. In a system under fault tolerant control (FTC), whenever a fault is detected, the fault amplitude is estimated in order to compensate its effect through an appropriate reconfiguration of the controller module [6, 7, 8]. The performance of the FTC system depends mainly on the estimation accuracy of the fault magnitude. As for fault detection, it is desirable that the fault estimation is robust with respect to noises and unexpected uncertainties and perturbations.

Most fault estimation approaches are optimisation-based, and thus optimisation techniques are used to solve the fault estimation problem. The interested reader is referred to [9]. This chapter looks into the problem of estimating faults using the proposed PCA-based KL divergence approach. The fault estimation accuracy will be defined in a probabilistic framework. There are two main objectives: first, derive an analytical model of KL divergence, which is specifically aimed at estimating incipient faults amplitude; second, evaluate the estimation accuracy of incipient faults in noisy environments. The next section presents assumptions made to the data and fault modelling in order to obtain the divergence model. Section 3 is dedicated to the analytical model derivation. In section 4, the fault amplitude will be expressed as function of the divergence and the PCA's parameters. Section 5 is

devoted to the validation of the obtained model and the evaluation of the theoretical fault estimation using a simulated AR process. In section 6, the divergence-based fault estimation using PCA to model multivariate data will be compared to the maximum likelihood estimate using minimum mean square estimation. The same AR process will be used to this end.

3.2 Assumptions

3.2.1 Data modelling

The divergence may be reduced to a simple and closed expression, when prior knowledge or assumptions about the characteristics of pdfs are available. Under the PCA framework the distributions along the original axes can be assumed as normal, because basically PCA yields an optimal representation for multivariate data which are at least approximately normal [10]. For these data, the principal subspace is spanned by the first l eigenvectors of the sample covariance/correlation matrix. Then the principal component scores, which are linear combinations of the original variables, can be assumed as being normally distributed along their axes. The first latent scores have, contrary to the residual ones, large variances so that their distributions are far from being degenerated. Therefore, the divergence is strongly related to the principal subspace, and the first latent principal scores will be considered to have normal distributions.

The process variables are affected with independent and identically distributed (i.i.d) Gaussian noise representing measurement errors. So the noise samples are considered to be drawn from a normal distribution with zero mean and variance σ_v^2. The noise added to each observation of m variables is denoted $\boldsymbol{v}$, such that $\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2 I_m)$ where I_m is the $(m \times m)$ identity matrix. Let $\boldsymbol{V}_{[N\times 1]}$ be a noise vector of N samples drawn from the distribution $\mathcal{N}(0, \sigma_v^2)$.

3.2.2 Fault modelling

The simple fault case is considered, that is when $\boldsymbol{x_j}$ is faulty, $\boldsymbol{x_r}$ with $r \neq j$ is fault-free. The fault that affects $\boldsymbol{x_j}$ is modelled such that its signature is incipient according to the signal properties. It is characterised with a multiplicative factor with amplitude a that affects the last $(N-b)/N$ observations of a signal acquisition. As it is shown in Fig.3.1, this fault model approaches an incipient fault, provided that $(N-b)/N$ and a are sufficiently low. It assumes that during the first stage of the incipient fault development, the fault amplitude (size or severity) is constant. Such a fault will not modify the centre and direction of the principal subspace, which

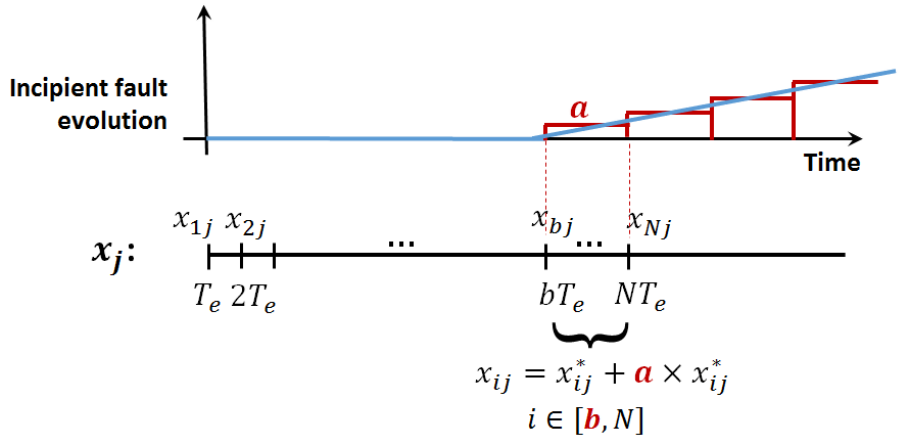

Figure 3.1: Incipient fault model

is thus supposed unchanged after the fault occurrence. The difficulty however is to detect the incipient fault in this stage where a significant loss of process operation performance can occur without being noticed. We may write for N measurements collected from the variable $\boldsymbol{x_j}$:

$$\boldsymbol{x_j} = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{bj} \\ \vdots \\ x_{Nj} \end{bmatrix} = \begin{bmatrix} x^*_{1j} \\ x^*_{2j} \\ \vdots \\ x^*_{bj} \\ \vdots \\ x^*_{Nj} \end{bmatrix} + a \times \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x^*_{bj} \\ \vdots \\ x^*_{Nj} \end{bmatrix} + \boldsymbol{V} = \boldsymbol{x^*_j} + \boldsymbol{F_j} + \boldsymbol{V} \tag{3.1}$$

where $\boldsymbol{F_j} = a \times [\ 0 \quad 0 \quad \dots \quad x^*_{bj} \quad \dots \quad x^*_{Nj}\ ]'$.

3.2.3 Assumptions's consequence

As a consequence, the covariance/correlation matrix S estimated from X is:

$$S = P^* \Lambda P^{*\prime} + \sigma_v^2 I_m \tag{3.2}$$

where

$$\Lambda = \Lambda^* + \Delta\Lambda \tag{3.3}$$

$\Lambda^* = diag(\lambda^*_1, ..., \lambda^*_l, 0, .., 0)$ is the $(m \times m)$ matrix of eigenvalues associated to eigenvectors $\boldsymbol{p^*_1}, ..., \boldsymbol{p^*_l}, \boldsymbol{p^*_{l+1}}, ..., \boldsymbol{p^*_m}$.
$\Delta\Lambda = diag(\Delta\lambda_1, ..., \Delta\lambda_m)$ is the change due to the fault occurrence. So, $\Delta\Lambda = 0$ when $a = 0$. The last $(m - l)$ eigenvalues correspond to the residual subspace. As

the corresponding distributions may vanish ($\lambda_k^* = 0$ for $k > l$), the divergence is concerned with only the first l principal scores for which $\lambda_k^* \neq 0$ ($k = 1, ..., l$).

Besides, given two normal densities h and q such that $h \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $q \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where μ_1, μ_2 are the means and σ_1^2, σ_2^2 are the variances for h and q respectively, the KL divergence between h and q may be written as

$$D_{An}(h, q) = \frac{1}{2}[\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + (\mu_1 - \mu_2)^2(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}) - 2]. \tag{3.4}$$

In the following, this expression will be used for computing the analytical PCA-based KL divergence model.

3.3 Analytical model derivation

3.3.1 KL divergence expression

From the normality assumption, it follows that each of the first l principal component scores $\boldsymbol{t_k}$, $k = 1, 2, ..., l$, has a pdf which we denote f_k such that $f_k \sim \mathcal{N}(0, \lambda_k + \sigma_v^2)$. λ_k is hence the variance of $\boldsymbol{t_k}$ and the eigenvalue associated to the kth principal component as well. We propose to compare f_k against its reference, which corresponds to the fault-free operating mode. The reference is denoted f_k^{rf}, $f_k^{rf} \sim \mathcal{N}(0, \lambda_k^* + \sigma_v^2)$. The mean of the distribution is supposed unchanged (zero) after the occurrence of a fault, because it has been assumed that a fault, particularly an incipient one, will not move the centre of the PCA's model. So it follows from Eq.3.3 that

$$\lambda_k = \lambda_k^* + \Delta\lambda_k \tag{3.5}$$

where $\Delta\lambda_k$ is the eigenvalue bias caused by the fault occurrence. By specialising Eq.3.4 to the case considered, the divergence between f_k and f_k^{rf} becomes:

$$D_{An}(f_k, f_k^{rf}) = \frac{1}{2}[\frac{\Delta\lambda_k^2}{(\lambda_k^* + \sigma_v^2)(\lambda_k^* + \sigma_v^2 + \Delta\lambda_k)}]. \tag{3.6}$$

The next step is to write $\Delta\lambda_k$ in function of the fault 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 $a \approx 0$, the Taylor development of λ_k gives

$$\lambda_k = \lambda_k^* + \frac{\partial\lambda_k}{\partial a}(0)a + \frac{1}{2}\frac{\partial^2\lambda_k}{\partial a^2}(0)a^2 + \frac{1}{3!}\frac{\partial^3\lambda_k}{\partial a^3}(0)a^3 + ... \tag{3.7}$$

Subsequently, in order to obtain the divergence expression depending on a, the eigenvalue-derivatives with respect to this parameter must be computed. Suppose

PCA is computed with the covariance/correlation matrix S, it is shown in [11] that writing S in function of a gives the nth-order eigenvalue derivative as

$$\frac{\partial^n \lambda_k}{\partial a^n} = \boldsymbol{p_k^*}' \frac{\partial^n S}{\partial a^n} \boldsymbol{p_k^*} \tag{3.8}$$

where $\boldsymbol{p_k^*}$ is the kth loading vector/eigenvector associated to λ_k^*. Similarly to λ_k^*, $\boldsymbol{p_k^*}$ refers to the fault-free PCA's model.

Thus, each element of S will be differentiated n times with respect to the fault parameter a. Both of the PCA cases, where it is applied to the covariance matrix and the correlation matrix of data, will be considered in the sequel.

3.3.2 Covariance matrix case

S is given by

$$S = \frac{1}{N} \bar{X}' \bar{X}. \tag{3.9}$$

which is an unbiased estimate of the true covariance matrix in case of multinormally distributed data. Thus,

$$S = \frac{1}{N} \begin{bmatrix} \boldsymbol{\bar{x}_1'}\boldsymbol{\bar{x}_1} & \dots & \boldsymbol{\bar{x}_1'}\boldsymbol{\bar{x}_j} & \dots & \boldsymbol{\bar{x}_1'}\boldsymbol{\bar{x}_m} \\ \vdots & & \vdots & & \vdots \\ \boldsymbol{\bar{x}_j'}\boldsymbol{\bar{x}_1} & \dots & \boldsymbol{\bar{x}_j'}\boldsymbol{\bar{x}_j} & \dots & \boldsymbol{\bar{x}_j'}\boldsymbol{\bar{x}_m} \\ \vdots & & \vdots & & \vdots \\ \boldsymbol{\bar{x}_m'}\boldsymbol{\bar{x}_1} & \dots & \boldsymbol{\bar{x}_m'}\boldsymbol{\bar{x}_j} & \dots & \boldsymbol{\bar{x}_m'}\boldsymbol{\bar{x}_m} \end{bmatrix}. \tag{3.10}$$

It follows from the fault model in Eq.3.1 that

$$\begin{aligned} \boldsymbol{\bar{x}_j} &= \boldsymbol{x_j} - \mu_j \mathbb{1} \\ &= (\boldsymbol{x_j^*} - \mu_j^* \mathbb{1}) + (\boldsymbol{F_j} - a \times \tfrac{1}{N} \textstyle\sum_{i=b}^{N} x_{ij}^* \mathbb{1}) + \boldsymbol{V} \\ &= \boldsymbol{\bar{x}_j^*} + \boldsymbol{\bar{F}_j} + \boldsymbol{V} \end{aligned} \tag{3.11}$$

where $\boldsymbol{\bar{F}_j} = \boldsymbol{F_j} - a \times \frac{1}{N} \sum_{i=b}^{N} x_{ij}^* \mathbb{1}$, $\mathbb{1}$ is a column vector of N ones and μ_j is the sample mean of $\boldsymbol{x_j}$. Based on Eq.3.2, the derivation of S with respect to a can be made while considering $\boldsymbol{V} = 0$. Substituting $\boldsymbol{\bar{x}_j}$ in S by its expression (3.11) gives after all calculations:

$$\frac{\partial \boldsymbol{\bar{x}_r'}\boldsymbol{\bar{x}_q}}{\partial a} = 0, \quad \forall r, q \neq j \tag{3.12}$$

$$\frac{\partial \boldsymbol{\bar{x}_r'}(\boldsymbol{\bar{x}_j^*} + \boldsymbol{\bar{F}_j})}{\partial a} = \frac{\partial (\boldsymbol{\bar{x}_j^*} + \boldsymbol{\bar{F}_j})'\boldsymbol{\bar{x}_r}}{\partial a} = \delta_r \quad \forall r \neq j \tag{3.13}$$

$$\frac{\partial(\bar{\boldsymbol{x}}_{\boldsymbol{j}}^{*}+\bar{\boldsymbol{F}}_{\boldsymbol{j}})'(\bar{\boldsymbol{x}}_{\boldsymbol{j}}^{*}+\bar{\boldsymbol{F}}_{\boldsymbol{j}})}{\partial a}=2\delta_j+2a\tau. \tag{3.14}$$

where δ_r, δ_j and τ are constants independent of the fault amplitude. They are given by:

$$\delta_r=\sum_{i=b}^{N}(x_{ir}^{*}-\mu_r^{*})x_{ij}^{*} \quad \forall r \tag{3.15}$$

$$\tau=\sum_{q=b}^{N}\left(x_{qj}^{*}-\frac{1}{N}\sum_{i=b}^{N}x_{ij}^{*}\right)^2 \tag{3.16}$$

They are functions of the original variables and can be computed from healthy data once for all. The first-order derivative of the covariance matrix is then given by:

$$\frac{\partial S}{\partial a}=\frac{1}{N}\begin{bmatrix} 0 & \dots & \delta_1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \delta_1 & \dots & 2\delta_j+2a\tau & \dots & \delta_m \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & \delta_m & \dots & 0 \end{bmatrix} \tag{3.17}$$

The second-order sensitivity of S with respect to the fault amplitude a is obtained by differentiating (3.17). It gives:

$$\frac{\partial^2 S}{\partial a^2}=\frac{1}{N}\begin{bmatrix} 0 & \dots & 0 \\ & \vdots & \\ \vdots & 2\tau & \vdots \\ & \vdots & \\ 0 & \dots & 0 \end{bmatrix} \tag{3.18}$$

The higher-order sensitivities of S ($n>2$) are all null, as for the eigenvalue derivatives.

Writing the loading vector $\boldsymbol{p}_{\boldsymbol{k}}^{*}$ as $\boldsymbol{p}_{\boldsymbol{k}}^{*}=\begin{bmatrix} p_{1k} & \cdots & p_{mk} \end{bmatrix}'$ it follows that

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

and from Eq.3.7,

$$\Delta\lambda_k=\lambda_k-\lambda_k^{*}=\frac{2}{N}p_{jk}\sum_{r=1}^{m}p_{rk}\delta_r\times a+\frac{1}{N}p_{jk}^2\tau\times a^2. \tag{3.19}$$

Based on Eq.3.6 and Eq.3.19, the analytical expression of the divergence between the pdf of the kth principal score and its reference is:

$$D_{An}(f_k, f_k^{rf}) = \frac{2}{N^2} \frac{\left(p_{jk} \sum_{r=1}^{m} p_{rk}\delta_r \times a + 1/2 p_{jk}^2 \tau \times a^2\right)^2}{(\lambda_k^* + \sigma_v^2)\left(\lambda_k^* + \sigma_v^2 + \frac{2}{N}\left(p_{jk} \sum_{r=1}^{m} p_{rk}\delta_r\right) \times a + \frac{1}{N} p_{jk}^2 \tau \times a^2\right)}. \tag{3.20}$$

This is a closed form expression that explicitly depends on the PCA's model parameters, the fault amplitude and other constant values calculated from fault-free measurements.

3.3.3 Correlation matrix case

The correlation coefficient between the rth and jth original variable/feature is given by

$$\begin{aligned} Corr(\boldsymbol{x_r}, \boldsymbol{x_j}) &= \frac{Cov(\boldsymbol{x_r}, \boldsymbol{x_j})}{\sqrt{Var(\boldsymbol{x_r})}\sqrt{Var(\boldsymbol{x_j})}} \\ &= \frac{\bar{\boldsymbol{x}}_{\boldsymbol{r}}' \bar{\boldsymbol{x}}_{\boldsymbol{j}}}{\sqrt{\bar{\boldsymbol{x}}_{\boldsymbol{r}}' \bar{\boldsymbol{x}}_{\boldsymbol{r}}}\sqrt{\bar{\boldsymbol{x}}_{\boldsymbol{j}}' \bar{\boldsymbol{x}}_{\boldsymbol{j}}}} \end{aligned} \tag{3.21}$$

where Cov is the covariance coefficient, and Var denotes the variance. Making its derivative and using

$$\begin{cases} \dfrac{\partial \bar{\boldsymbol{x}}_{\boldsymbol{r}}{}' \bar{\boldsymbol{x}}_{\boldsymbol{j}}}{\partial a} = \delta_r \\ \dfrac{\partial \bar{\boldsymbol{x}}_{\boldsymbol{j}}{}' \bar{\boldsymbol{x}}_{\boldsymbol{j}}}{\partial a} = 2(\delta_j + a\tau) \end{cases}$$

give

$$\frac{\partial Corr(\boldsymbol{x_r}, \boldsymbol{x_j})}{\partial a} = \frac{\delta_r Var(\boldsymbol{x_j}) - (\delta_j + a\sigma)\, Cov(\boldsymbol{x_r}, \boldsymbol{x_j})}{\sqrt{Var(\boldsymbol{x_r})}\sqrt{(Var(\boldsymbol{x_j}))^3}}. \tag{3.22}$$

The calculation of the second-order derivative gives

$$\begin{aligned} \frac{\partial^2 Corr(\boldsymbol{x_r}, \boldsymbol{x_j})}{\partial a^2} = \frac{1}{N} &\frac{\left[\delta_r(\delta_j + a\sigma) - N\sigma Cov(\boldsymbol{x_r}, \boldsymbol{x_j})\right] Var(\boldsymbol{x_j})}{\sqrt{Var(\boldsymbol{x_r})}\sqrt{(Var(\boldsymbol{x_j}))^5}} - \\ &\frac{3}{N} \frac{\left[\delta_r Var(\boldsymbol{x_j}) - (\delta_j + a\sigma) Cov(\boldsymbol{x_r}, \boldsymbol{x_j})\right](\delta_j + a\sigma)}{\sqrt{Var(\boldsymbol{x_r})}\sqrt{(Var(\boldsymbol{x_j}))^5}}. \end{aligned} \tag{3.23}$$

As it is can be seen from the above equations, correlation coefficients are non-rational functions of the fault amplitude due to the variance term $\sqrt{(Var(\boldsymbol{x_j}))^3}$ in the denominator. The derivative expressions are more complicated than those obtained with the covariance matrix case, and also allow only an approximation of the eigenvalue bias because the non-zero derivatives make Taylor series infinite. We may write for the kth eigenvalue of the data correlation matrix:

$$\lambda_k = \lambda_k^* + \frac{\partial \lambda_k}{\partial a}(0)a + \frac{1}{2}\frac{\partial^2 \lambda_k}{\partial a^2}(0)a^2 + ... \tag{3.24}$$

with

$$\frac{\partial^n \lambda_k}{\partial a^n} = 2\sum_{r=1}^{m} p_{jk}p_{rk}\frac{\partial^n Corr(\boldsymbol{x_r}, \boldsymbol{x_j})}{\partial a^n}. \tag{3.25}$$

Then taking the first h eigenvalue derivatives, the divergence expression becomes

$$D_{An}(f_k, f_k^{rf}) = \frac{\left(\sum_{n=1}^{h} \frac{1}{n!}\frac{\partial^n \lambda_k}{\partial a^n}a^n\right)^2}{\lambda_k^*(\lambda_k^* + \sum_{n=1}^{h} \frac{1}{n!}\frac{\partial^n \lambda_k}{\partial a^n}a^n)}. \tag{3.26}$$

The correlation matrix gives a complex divergence expression. Taking only the first h derivatives supposes that the remaining terms are negligible. (3.26) is not a rational function of the fault amplitude a due to the function $\sqrt{(Var(\boldsymbol{x_j}))^3}$ in (3.22).

3.4 Fault estimation

3.4.1 Fault amplitude estimator

(3.20) and (3.26) are the analytical expressions of the divergence applied to the principal components of the covariance matrix and the correlation matrix respectively. The covariance matrix however yields a divergence model, which is a closed and rational function of the fault amplitude. (3.20) is suitable to derive a fault amplitude estimate that depends on the divergence value.

Let $\alpha_1 = p_{jk}\sum_{r=1}^{m} p_{rk}\delta_r$ and $\alpha_2 = 1/2p_{jk}^2\tau$. (3.20) gives the theoretical estimation of the fault amplitude [12] as:

$$\hat{a} = \frac{-\alpha_1 + \sqrt{\alpha_1^2 + 2\alpha_2 N(\lambda_k^* + \sigma_v^2)(\hat{D} + \sqrt{(\hat{D}^2 + 2\hat{D})})}}{2\alpha_2} \tag{3.27}$$

In practice, $\hat{D}$ is the divergence value computed using (2.9). Since (2.9) is just an approximation of the true value, the accuracy of estimating a through (3.27) should be evaluated.

3.4.2 Probabilistic model for the fault amplitude estimate

If a known distribution fits to the variable $\hat{D}$ computed through (2.9), the distribution of $\hat{a}$ can be obtained from the following theorem [13]:
Let $\boldsymbol{X}$ have pdf $f_{\boldsymbol{X}}(x)$ and let $\boldsymbol{Y} = g(\boldsymbol{X})$, where g is a monotone function. Let $\mathcal{X} = \{x : f_{\boldsymbol{X}}(x) > 0\}$ and $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$. Suppose that $f_{\boldsymbol{X}}(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivation on $\mathcal{Y}$, then the pdf of $\boldsymbol{Y}$ is given by:

$$f_{\boldsymbol{Y}}(y) = \begin{cases} f_{\boldsymbol{X}}(g^{-1}(y)) \mid \dfrac{d}{dy} g^{-1}(y) \mid & y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases} \tag{3.28}$$

So, Consider

$$g(x) = \frac{-\alpha_1 + \sqrt{\alpha_1^2 + N\alpha_2(\lambda_k^* + \sigma_v^2)(x + \sqrt{(x^2 + 2x)})}}{\alpha_2} \tag{3.29}$$

where the variable x refers to $\hat{D}$. The calculation of $g'(x)$ proves that g is monotone ($g'(x) > 0 \; \forall \; x \geq 0$). The inverse function of g is:

$$g^{-1}(y) = \frac{1}{2} \frac{(\beta_1 y + \beta_2 y^2)^2}{\beta(\beta + \beta_1 y + \beta_2 y^2)}, \tag{3.30}$$

where the variable y refers to $\hat{a}$, $\beta = \lambda_k^* + \sigma_v^2$, $\beta_1 = 2\alpha_1/N$ and $\beta_2 = \alpha_2/N$.
The derivation of $g^{-1}(y)$ is:

$$\left(g^{-1}(y)\right)' = \frac{4\beta_1\beta_2^2 y^4 + 2(\beta_1^2\beta_2 + \beta\beta_2^2)y^3 + 3\beta\beta_1\beta_2 y^2 + 2\beta\beta_1^2 y}{\beta(\beta + \beta_1 y + \beta_2 y^2)^2}. \tag{3.31}$$

Numerical simulations yield that the probability density $f_{\boldsymbol{X}}(\hat{D})$ fits with Gamma distributions defined with two parameters, known as shape α and scale θ parameters. It follows that

$$f_{\boldsymbol{X}}(\hat{D}) \approx \frac{\hat{D}^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} \exp(-\hat{D}/\theta). \tag{3.32}$$

where $\Gamma(\alpha)$ is the gamma function evaluated at α. According to (3.28) the pdf of $\hat{a}$ is:

$$f_{\boldsymbol{Y}}(\hat{a}) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \left(g^{-1}(\hat{a})\right)' \left(g^{-1}(\hat{a})\right)^{\alpha-1} \exp\left(\frac{-g^{-1}(\hat{a})}{\theta}\right) \tag{3.33}$$

α and θ are estimated from $\hat{D}$ computed on many realisations (see Fig.2.5).

3.5 Simulation results and discussions

3.5.1 Procedure description

The numerical example in chapter 2 is used for its nonlinear functions that make the probability distributions of generated signals arbitrary. This is relevant for proving the detection capability of the divergence. For estimation purposes, a distribution assumption has to be made to obtain an analytical model, and then followed in the evaluation stage. Therefore, we here opt for a multivariate AR system, in which the generated signals are approximately normally distributed. It is inspired from [14] and defined as follows:

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

$$y(i) = x(i) + v(i) \tag{3.34}$$

where u is the correlated input,

$$u(i) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix} u(i-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} w(i-1). \tag{3.35}$$

w is a vector of 2 inputs $w = [w_1 \quad w_2]'$, which are uncorrelated Gaussian signals with zero mean and unit variance. $u = [u_1 \quad u_2]'$ is the vector of measured inputs, and $y = [y_1 \quad y_2]'$ is the vector of outputs corrupted by uncorrelated Gaussian errors $v = [v_1 \quad v_2]'$ with zero mean and variance σ_v^2.

The vector of process variables will be formed with the measured inputs and outputs of the process at instant i, i.e. $[y_1(i) \quad y_2(i) \quad u_1(i) \quad u_2(i)]'$. A data matrix X of N measurements/rows is formed with these variables. PCA is applied on the corresponding covariance matrix; it gives 4 principal components with variances $\lambda_k = \{40.26, 4.9, 1.14, 0.17\}$, $k = 1, 2, 3, 4$. The first principal component accounts for 86.6% of variation, along with the second principal component they account for 97% of variation. The divergence will be evaluated on $\boldsymbol{t_k} = \bar{X}\boldsymbol{p_k^*}$, $k = 1, 2$, for various SNR (various σ_v^2).

Let $\hat{f}_k^{rf}$ and $\hat{f}_k$ be the normalised Gaussian kernel estimate of probability densities of $\boldsymbol{t_k}$ before and after the fault respectively. The divergence between $\hat{f}_k^{rf}$ and $\hat{f}_k$ is approximated as:

$$\hat{D}_{t_k} = \hat{D}(\hat{f}_k, \hat{f}_k^{rf}) = \sum_{t_k} (\hat{f}_k - \hat{f}_k^{rf}) \log \left[\frac{\hat{f}_k}{\hat{f}_k^{rf}} \right]. \tag{3.36}$$

A fault can affect one of the variables or a matrix coefficient. As to be consistent with the fault modelling adopted here, we consider the fault that affects the outputs

and can be seen as sensor fault. So the output vector is

$$\begin{bmatrix} y_1(i) \\ y_2(i) \end{bmatrix} = \begin{bmatrix} 1+a_1 & 0 \\ 0 & 1+a_2 \end{bmatrix} \begin{bmatrix} x_1(i) \\ x_2(i) \end{bmatrix} + \begin{bmatrix} v_1(i) \\ v_2(i) \end{bmatrix} \tag{3.37}$$

where a_1 and a_2 are the fault amplitudes on y_1 and y_2 respectively. y_1 is supposed to be affected by the fault, and thus $a_1 \neq 0$ in the sampling interval $[b, N]$. The simple fault case is considered and thus $a_2 = 0 \ \forall i$.

In the following, model validation results are shown. The numerical divergence (3.36) is compared to the analytical model (3.20), and then the probabilistic model of the estimated fault amplitude is depicted. Afterwards, the accuracy of the theoretical estimated fault amplitude is evaluated by computing the relative error on the estimation of the faulty variable. The relative errors are computed in different conditions showing the impact of the noise level (SNR), the proportion of the faulty sample ($(N-b)/N$), and the sample size (N) on the estimation accuracy.

3.5.2 Model validation

3.5.2.1 Divergence model

In the model validation, a large sample generated from the process is considered in order to minimize the error due to the divergence approximation [15]. We propose here to consider $N = 10^6$, which can be obtained in practice with, for example, a 10 kHz sampling frequency during 1mn 40 sec. To illustrate the divergence evolution with respect to the fault amplitude we consider the variation range $[0, 0.3]$ for a_1, meaning $[0, 30]\%$ variation of y_1's amplitude. The last 10% of the considered observations for y_1 are supposed affected by the fault. The proportion of the faulty interval to the total sample size is thus $(N-b)/N = 0.1$. The divergence is evaluated on the first and the second principal component $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ for various values of signal-to-noise ratio (SNR). The constants δ_r $(r = 1, ..., 4)$ and τ are calculated once for all from healthy data.

The analytical model will not be sensitive to the SNR value because it is obtained on the basis of the same noise added to the reference and healthy data. Accordingly, the theoretical divergence will only depend on the fault amplitude whatever the noise variance. Fig.3.2 and Fig.3.3 display $\hat{D}$ computed on $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ versus the fault amplitude a_1. $D_{An}(f_k, f_k^{rf})$ refers to the analytical model (3.20) and the other lines are obtained from the divergence approximation (3.36). In Fig.3.2, the theoretical model evaluated on $\boldsymbol{t_1}$ $(D_{An}(f_1, f_1^{rf}))$ shows an important gap with the approximated divergence especially at low values of a_1. Low noise levels, high SNR, are considered yet. However, the model once calculated on $\boldsymbol{t_2}$ $(D_{An}(f_2, f_2^{rf}))$,

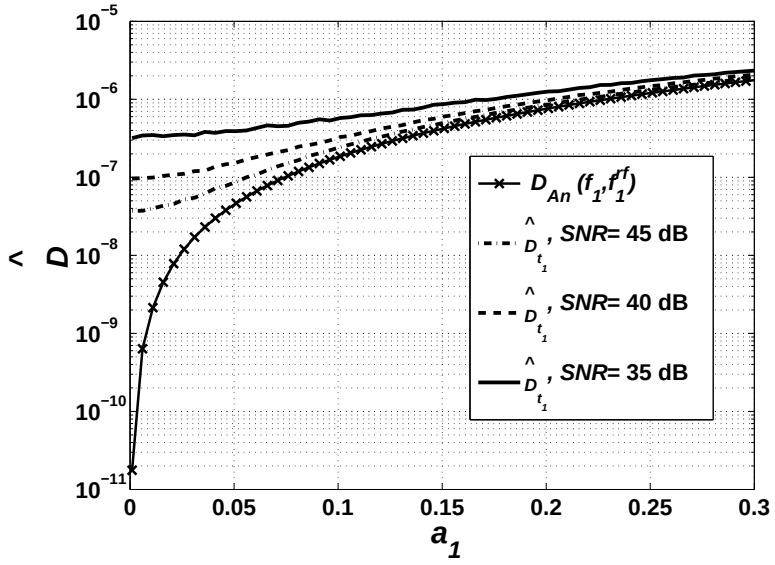

Figure 3.2: Evolution of KL divergence computed on $\boldsymbol{t_1}$

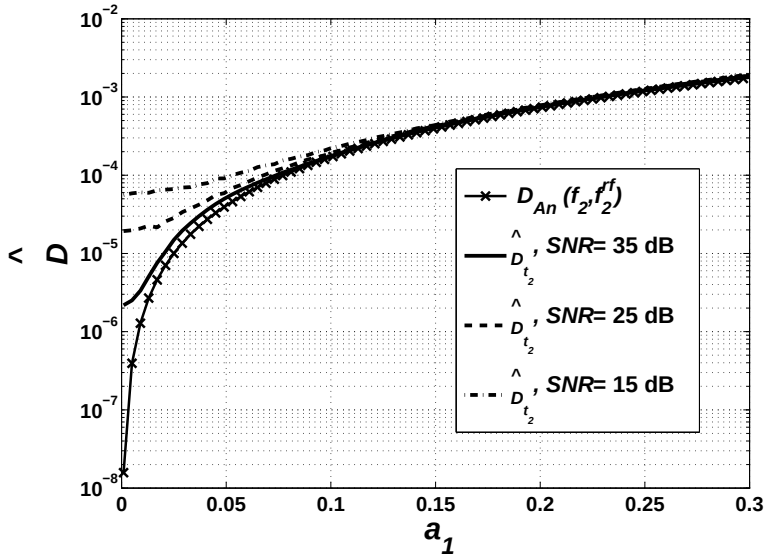

Figure 3.3: Evolution of the KL divergence computed on t_2

fits better with the approximated divergence ($\hat{D}_{t_2}$) as shown in Fig.3.3. The gap increases with decreasing SNR, but it is still relatively negligible especially for high fault amplitudes, $a_1 > 0.1$. To explain these results we must refer to the loading vector $\boldsymbol{p_k^*}$ that spans each of the principal component variable [12].

The coefficients of loading vectors displayed in Table 3.1 show the rate of the

process variables (y_1, y_2, u_1 and u_2) contribution to each of the principal component. The faulty variable y_1 contributes with a high rate to $\boldsymbol{t_2}$ (0.93) compared

Table 3.1: Loading vectors coefficients

	$\boldsymbol{p_1}$	$\boldsymbol{p_2}$
p_{1k}	-0.12	0.93
p_{2k}	0.97	0.08
p_{3k}	0.16	0.11
p_{4k}	0.04	0.324

to its contribution to $\boldsymbol{t_1}$ (0.12). The divergence computed on $\boldsymbol{t_2}$ will be thus more sensitive than $\boldsymbol{t_1}$ to the small amplitude shifts that affect y_1. Accordingly, a better result of model fit will be obtained on $\boldsymbol{t_1}$ if the fault affects y_2 (p_{21}=0.97) or u_1 (p_{31}=0.16) instead of y_1. Whatever the case, the error between the analytical and the approximated divergence normally decreases with increasing SNR.

It stems from the previous results that detecting and estimating a fault that affects the output y_1 will be based on the divergence computed on $\boldsymbol{t_2}$. The results of estimation will therefore be shown for this case. The worst case, when the fault estimation is carried out on the unfavourable principal component, was studied in [12].

3.5.2.2 Fault amplitude estimation model

The divergence calculated in particular conditions of fault and noise through (2.9), is a random variable whose probability distribution can be estimated on a sample of $\hat{D}_{t_2}$ obtained through many realisations (Fig.2.5). Fig.3.4 shows the Gamma distribution fitting points that overlap the estimated distribution of $\hat{D}_{t_2}$ computed in case SNR= 35 dB, and for several fault amplitudes. $(N-b)/N$ is still set equal to 0.1. The Gamma distribution is a two-parameter distribution, which in special cases of these parameters, known as shape α and scale θ parameters, gives the exponential and chi-squared distributions. In particular, the distribution converges to a normal one for large shape parameter (as $\alpha \to \infty$).

Using the Gamma distribution parameters estimated from $\hat{D}$ samples, the probability distribution model obtained in (3.33) can be computed. Fig.3.5 depicts the (3.33) result corresponding to Fig.3.4. The estimated amplitude is still non-zero in fault-free condition ($a_1 = 0$), since the divergence value is non-zero in this case. Fig.3.5 shows that the estimator (3.27) gives an overestimation of the actual fault amplitude.

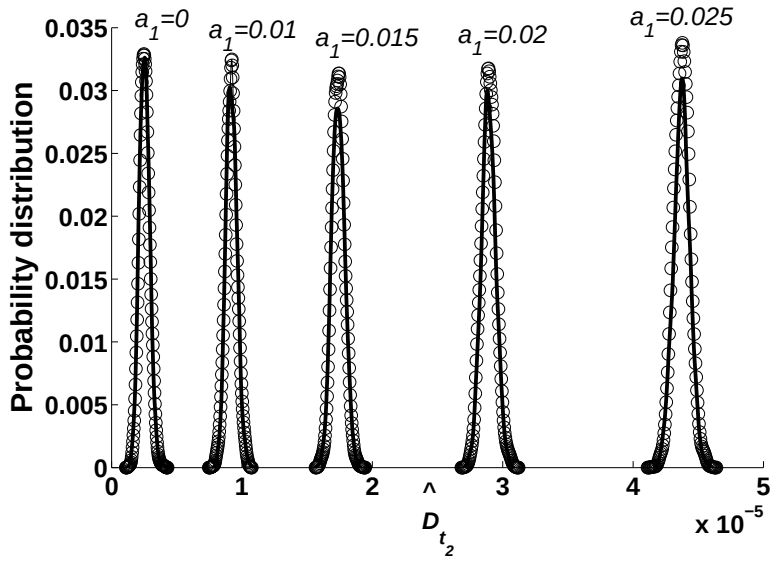

Figure 3.4: Gamma distribution fitting to probability distribution of $\hat{D}_{t_2}$

The estimation precision can be evaluated as being the reciprocal of the estimator

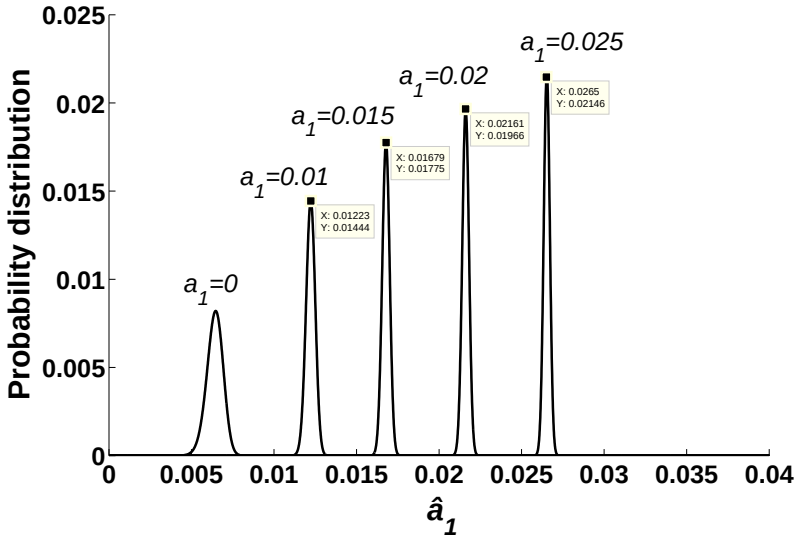

Figure 3.5: Probability distribution (3.33) of $\hat{a_1}$

standard deviation. Here, the precision order is about 10^3, meaning a high estimation accuracy. By taking the means of pdfs and considering a wide fault amplitude range that goes from 0.001 (0.1%) to 0.3 (30%), we obtain Fig.3.6 that plots the actual fault amplitude and the corresponding estimation mean. The error between the

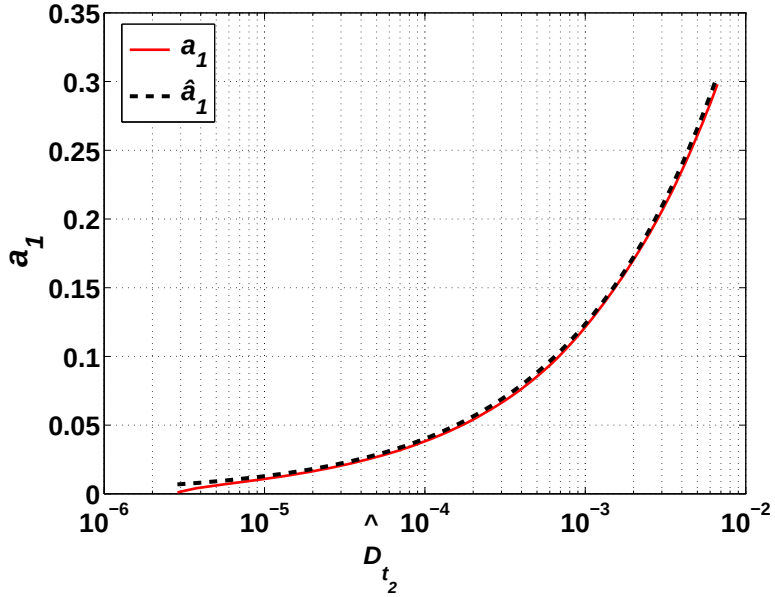

Figure 3.6: Actual and estimated fault amplitude

two curves will be assessed for different noise levels, data sizes and fault severities.

3.5.3 Estimation error

The objective of the fault estimation is often being able to recover the valuable information contained into the fault-free signal. Therefore, it seems relevant to assess the relative error on the estimation of the faulty signal, here y_1. We can write for the estimation of the faulty observation $y_1(i)$:

$$\hat{y}_1(i) = (1 + \hat{a}_1)x_1(i) + v_1(i) \tag{3.38}$$

The relative error, denoted E_r, on the estimation of y_1 is:

$$E_r = \frac{\hat{y}_1(i) - y_1(i)}{y_1(i)} \cong \frac{\hat{a}_1 - a_1}{1 + a_1}. \tag{3.39}$$

E_r will be evaluated in function of the fault amplitude, under different noise levels and different sizes of data and faulty sample.

3.5.3.1 Impact of faulty sample length

Fig.3.7, 3.8 and 3.9 show the relative error on the estimation of y_1 obtained at different SNR for a_1 ranging from 0.001 (0.1%) to 0.3 (30%). N is still set equal to 10^6. For particular a_1 and $(N - b)/N$ values, the error clearly increases with decreasing

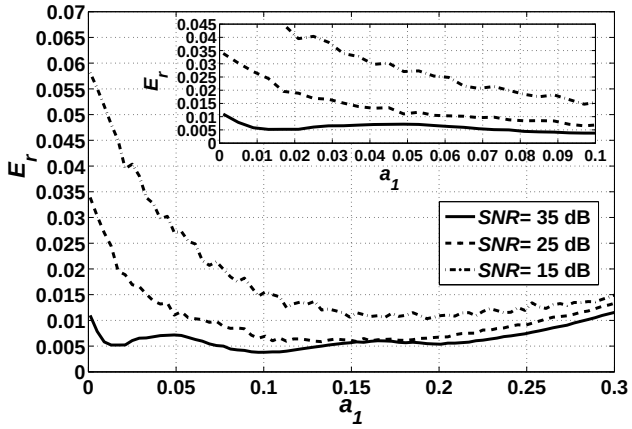

Figure 3.7: Estimation relative error E_r, $(N-b)/N = 0.1$

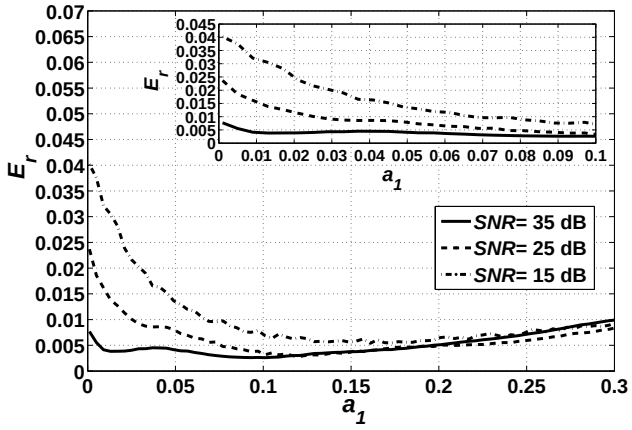

Figure 3.8: Estimation relative error E_r, $(N-b)/N = 0.15$

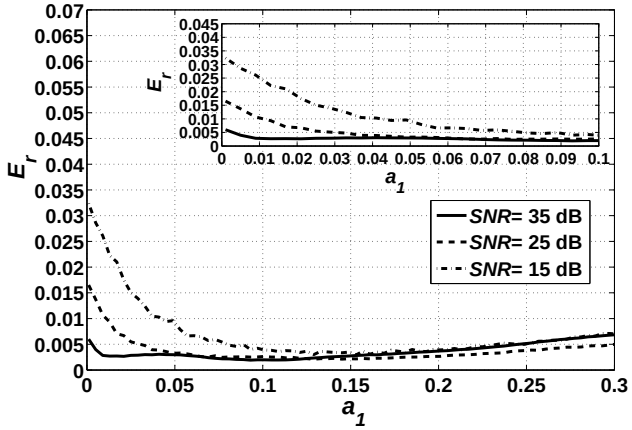

Figure 3.9: Estimation relative error E_r, $(N-b)/N = 0.2$

SNR. The figures show a zoom in the range of small amplitudes $a_1 < 0.1$. The relative error E_r is relatively high in this range. It decreases however with increasing proportion $(N-b)/N$ of faulty sample. It follows that in favourable conditions of noise and fault, say in case the SNR= 35 dB, the fault amplitude $a_1 > 0.1(10\%)$ and the proportion of faulty sample $(N-b)/N > 0.2$, the relative estimation error $E_r < 0.005$ (0.5%). In worst conditions, characterised with a high noise level (SNR= 15 dB) and incipient fault amplitude ($a_1 < 0.1$, $(N-b)/N = 0.1$), E_r reaches 0.06 (6%). E_r has nevertheless a positive value, which means an overestimation. In process monitoring, the overestimation of the fault amplitude/severity provides a safety margin for the process being monitored.

3.5.3.2 Impact of the data size

The results shown beforehand are obtained for a data size N= 10^6 observations. It is important however, from a practical point of view, to evaluate the estimation accuracy for smaller data sizes. Fig.3.10 illustrates the estimation error in the range of incipient fault amplitude for different data sizes N, SNR= 25 dB and $(N-b)/N$= 0.1. It shows that the accuracy of estimating the incipient fault amplitude using

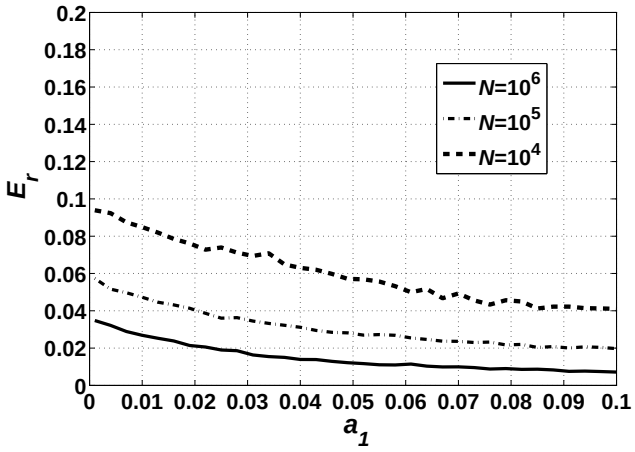

Figure 3.10: Estimation relative error E_r, $(N-b)/N = 0.1$, $SNR = 25$ dB

(3.27) depends strongly on the data size N, such that E_r readily increases with decreasing N. Fig.3.11 shows the same result for more favourable fault conditions, the proportion of faulty sample being set equal to 0.2. E_r is relatively lower than those obtained previously and can still be reduced by using a larger data sample.

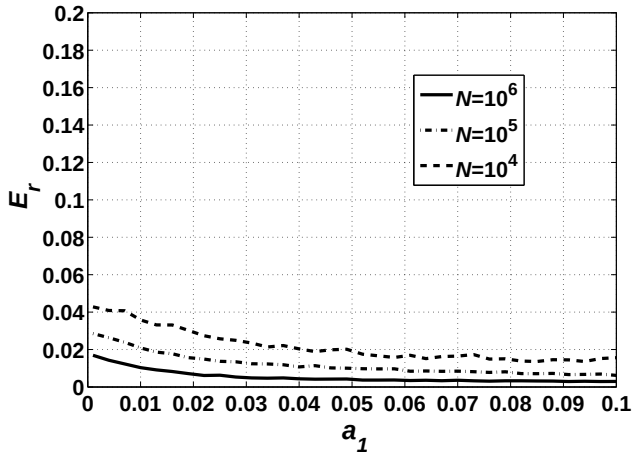

Figure 3.11: Estimation relative error E_r, $(N-b)/N = 0.2$, $SNR = 25$ dB

3.6 Comparison to another estimation method

The divergence-based fault estimation using PCA to model multivariate data, will be compared to the maximum likelihood estimate using minimum mean square estimation (MMSE). Assuming that the process variables are redundant and are simultaneously sampled, the MMSE allows estimating each variable using the others. The process variables are arranged into a vector $\mathbf{x}$:

$$\mathbf{x} = \begin{Bmatrix} \mathbf{u} \\ \mathbf{w} \end{Bmatrix} \tag{3.40}$$

where $\mathbf{u}$ is the faulty variable x_j, and $\mathbf{w}$ are the other observed variables. The covariance matrix $\mathbf{\Sigma}$ of $\mathbf{y}$, estimated using the training data, can be partitioned as:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{uu} & \mathbf{\Sigma}_{uw} \\ \mathbf{\Sigma}_{wu} & \mathbf{\Sigma}_{ww} \end{bmatrix} \tag{3.41}$$

The linear MMSE estimator $\hat{\mathbf{u}}$ for $\mathbf{u}|\mathbf{w}$ ($\mathbf{u}$ given $\mathbf{w}$) is obtained by minimising the mean-square error [16] $\epsilon_{MS} = \mathbb{E}((\mathbf{u}-\hat{\mathbf{u}})'(\mathbf{u}-\hat{\mathbf{u}}))$. The optimal estimator that minimises ϵ_{MS} is:

$$\hat{\mathbf{u}} \equiv E(\mathbf{u}|\mathbf{w}) = \mathbf{\Sigma}_{uw}\mathbf{\Sigma}_{ww}^{-1}\mathbf{w} \tag{3.42}$$

If x_j, here $\mathbf{u}$, is not faulty, the i_{th} sample of $\mathbf{u}$ is modelled as:

$$\mathbf{u}(i) = \hat{\mathbf{u}}(i) + \mathbf{v}(i) \tag{3.43}$$

where $\hat{\mathbf{u}}$ is given by (3.42) and $\mathbf{v}$ is the Gaussian noise with variance σ_v^2. A gain fault is modelled as:

$$\mathbf{u}(i) = \mathbf{G}\hat{\mathbf{u}}(i) + \mathbf{v}(i) \tag{3.44}$$

where $\mathbf{G} = 1 + a$. Assuming a Gaussian noise with zero-mean gives the conditional probability density function of $\mathbf{u}|\mathbf{w}$ as:

$$p(\mathbf{u}|\mathbf{w}) = p(\mathbf{u}(i)) = \frac{1}{\sqrt{2\pi\sigma_v^2}} exp\left[-\frac{1}{2\sigma_v^2}(\mathbf{u}(i) - \mathbf{G}\hat{\mathbf{u}}(i))^2\right] \tag{3.45}$$

in which σ_v^2 is obtained using the training data by

$$\sigma_v^2 = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{u}(i) - \hat{\mathbf{u}}(i))^2. \tag{3.46}$$

The maximum likelihood estimate (MLE) of $\mathbf{G}$ is obtained by maximising the log-likelihood [1]

$$log\ p(\mathbf{u}; \mathbf{G}) = -\frac{N}{2} log(2\pi\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{i=1}^{N}(\mathbf{u}(i) - \mathbf{G}\hat{\mathbf{u}}(i))^2 \tag{3.47}$$

Taking the derivative of the log-likelihood function with respect to $\mathbf{G}$ and setting the result equal to zero, produces

$$\hat{\mathbf{G}} = \frac{\sum_{i=1}^{N} \hat{\mathbf{u}}(i)\mathbf{u}(i)}{\sum_{i=1}^{N} \hat{\mathbf{u}}(i)^2} \tag{3.48}$$

This estimator assumes that the fault exists during the whole measurement. So the comparison with the divergence-based estimator will be done under this assumption, i.e. for $(N - b)/N = 100\%$. We evaluate the accuracy of both estimators using the following procedure.

1. N training samples are generated from the above AR process. The covariance matrix of $[y_1, y_2, u_1, u_2]$ is estimated.

   i. PCA is applied on the training covariance matrix to obtain the principal components parameters. δ_r and τ are also calculated.

   ii. $\mathbf{\Sigma}_{uw}$ and $\mathbf{\Sigma}_{ww}$ are extracted from the covariance matrix.

2. N samples of the data vector $[y_1, y_2, u_1, u_2]$ are generated with faulty y_1. $y_1(i) = (1 + a)y_1^*(i) + v_1(i)$, where $(1 + a) = \mathbf{G}$ and $i = 1, ..., N$.

   i. The divergence-based estimation is applied. a is estimated according to (3.27).

   ii. The maximum likelihood estimate of $\mathbf{G}$ is calculated according to (3.48).

3. The relative error of estimation is evaluated for both estimators:

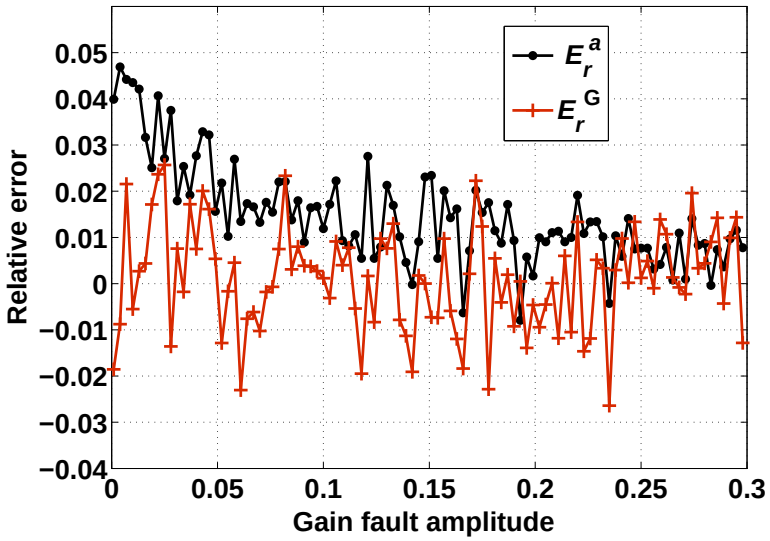

Figure 3.12: Estimation relative error

i. $E_r^a = (\hat{a} - a)/(1 + a)$.

ii. $E_r^{\mathbf{G}} = (\hat{\mathbf{G}} - \mathbf{G})/(\mathbf{G})$.

Step 1. is done once for all on training data. Step 2. and 3. are done for each gain fault value. The results will be shown for N=1000, SNR=15 dB and fault amplitude a ranging from 0 to 30%. Fig.3.12 shows the relative estimation error obtained while following the above procedure. It can be remarked that the standard deviation of $E_r^{\mathbf{G}}$, for the estimator $\hat{\mathbf{G}}$, is greater than that of E_r^a, for the estimator $\hat{a}$. Also, $E_r^{\mathbf{G}}$ is regularly negative, meaning that $\hat{\mathbf{G}}$ may underestimate the true fault amplitude. In this figure, it is also shown that E_r^a exceeds $E_r^{\mathbf{G}}$ in the small amplitude range ($a < 5\%$). In order to fairly compare the two estimators, we have to show the interval of variation of the relative error for each one. Consequently, the steps 2. and 3. are performed $R = 1000$ times at each fault value, and the variation intervals are calculated as $[\mu - 3\sigma, \mu + 3\sigma]$ where μ and σ are the average and the standard deviation of the obtained (1000×1) vectors of relative error. For example, at $a = 0.1$, equivalently $\mathbf{G}$=1.1, Fig.3.13 displays the histograms of E_r^a and $E_r^{\mathbf{G}}$. The histograms fit with normal distribution, which is the reason for using the mentioned interval. The result is shown in Fig.3.14, which displays the variation interval of the estimation error associated to each estimator. Dashed line is the calculated average. However, simulation showed that in unfavourable conditions of data size and noise (small data size N=1000, high noise level SNR=15 dB), the result of the MLE is not exactly reproducible. A slightly different MLE result can be obtained when using another training data generated at step. 1. There is some shift in the average of $E_r^{\mathbf{G}}$ but its standard deviation is still the same. An example is shown in Fig.3.15.

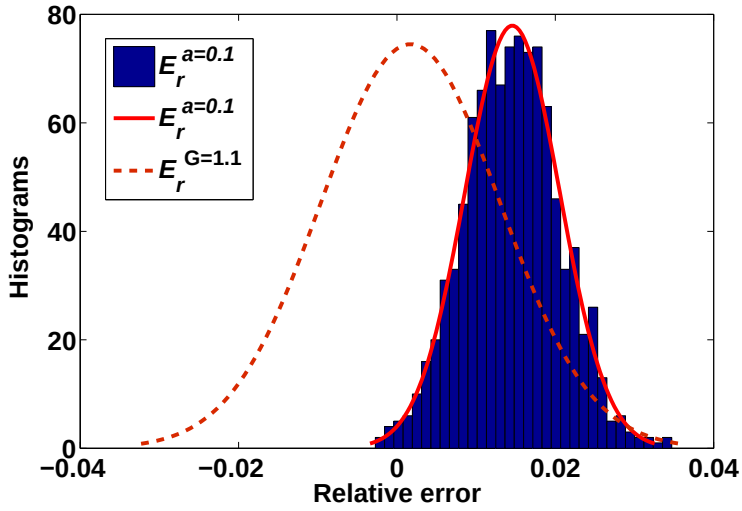

Figure 3.13: Relative error histograms

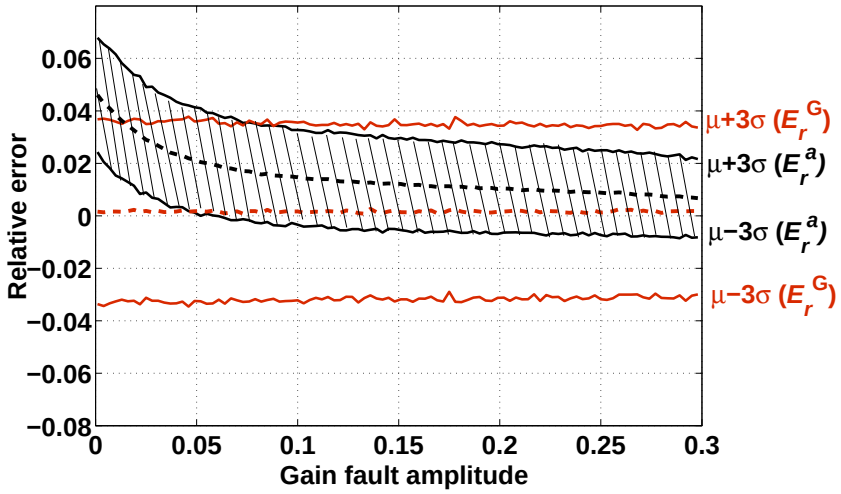

Figure 3.14: Estimation relative error interval

However, the divergence-based estimation using PCA is more robust to step 1 than the MLE using MMSE, since the associated result is quite reproducible. Moreover, E_r^a has a narrower variation range, which is mostly in the $E_r^{\mathbf{G}}$ range. It is mainly positive, meaning an overestimation of the fault amplitude.

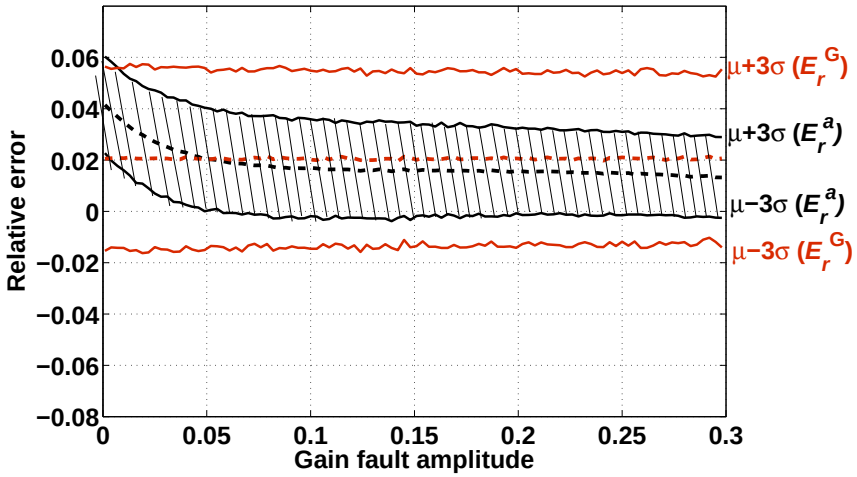

Figure 3.15: Estimation relative error interval

3.7 Conclusion

In PCA-based data representation, KL divergence has been proposed to estimate the incipient fault amplitude. A theoretical fault amplitude estimate that depends on the divergence value has been obtained. With numerically approximated divergence, the accuracy of the fault estimate has been evaluated while showing the impact of critical conditions of noise and data (high noise levels and low data sizes).

The divergence-based fault estimation using PCA to model multivariate data, has been compared to the maximum likelihood estimate using minimum mean square estimation. Our estimator is more robust to noisy data and has a lower standard deviation. The divergence-based estimator proved to be an overestimation (estimation error $E_r > 0$) of the actual fault amplitude, guaranteeing a safety margin for the monitored process.

The next chapter will focus on the application of the proposed tools on experimental data. Experimental data generated by real incipient faults will be used to show the effectiveness of the KL divergence and data-driven tools in the detection and diagnosis of small faults, whose signature is weak and concealed by the noise.

Bibliography

[1] J. Kullaa, "Detection, identification, and quantification of sensor fault in a sensor network," *Mechanical Systems and Signal Processing*, vol. 40, pp. 1208–221, 2013.

[2] S. Wang and J. Cui, "Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method," *Applied Energy*, vol. 82, pp. 197–213, 2005.

[3] S. A. Arogeti, D. Wang, C. B. Low, and M. Yu, "Fault detection isolation and estimation in a vehicle steering system," *IEEE Transactions on Industrial Electronics*, vol. 59, pp. 4810–4820, 2012.

[4] T. N. Pranatyasto and S. J. Qin, "Sensor validation and process fault diagnosis for fcc units under mpc feedback," *Control Engineering Practice*, vol. 9, pp. 877–888, 2001.

[5] Z. Gao and S. X. Ding, "Sensor fault reconstruction and sensor compensation for a class of nonlinear state-space systems via a descriptor system approach," *Control Theory & Applications, IET*, vol. 1, pp. 578–585, 2007.

[6] C. Zhaohui and H. Noura, "Sensor fault reconstruction and sensor compensation for a class of nonlinear state-space systems via a descriptor system approach," *8th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 236–241, 2013.

[7] M. Liu and P. Shi, "Sensor fault estimation and tolerant control for ito stochastic systems with a descriptor sliding mode approach," *Automatica*, vol. 49, pp. 1242–1250, 2013.

[8] Z. Gao and S. Ding, "Fault estimation and fault-tolerant control for descriptor systems via proportional, multiple-integral and derivative observer design," *Control Theory & Applications, IET*, vol. 1, pp. 1208 –1218, 2007.

[9] J. Stoustrup and H. H. Niemann, "Fault estimation - a standard problem approach," *International journal of robust and nonlinear control*, vol. 12, pp. 649–673, 2002.

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

[11] 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, 2007.

[12] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part II," *Elsevier Signal Processing*, 2014, accepted.

[13] G. Casella and R. L. Berger, *Statistical Inference*. Thomson Learning, 2002.

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

[15] J. Harmouche, C. Delpha, and D. Diallo, "A theoretical approach for incipient fault severity assessment using the Kullback-Leibler divergence," *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, Sept. 2013.

[16] J. Kullaa, "Sensor validation using minimum mean square error estimation," *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1444–1457, 2010.

Chapter 4

Application to non-destructive inspection-based damage detection

4.1 Introduction

This chapter is concerned with the application of the proposed data-driven tools to the detection, estimation and diagnosis of mechanical damages in engineering systems. A damage is here meant to denote a material crack which can be due to an external load (pressure, temperature, etc.), inevitable ageing, manufacturing imperfection, bad mounting, fatigue, and so on. The ability to rapidly assess the structural health of a system, without having to proceed to the disassembly of its elements is an important economic issue. This in-situ diagnostic capability would allow to decreasing maintenance costs while ensuring a sufficient security degree and a satisfactory operational level. The ability to identify small imperceptible cracks in a structure without altering its properties is possible by means of non-destructive inspection (NDI) or testing (NDT) techniques [1]. Several NDI methods have been applied to detect mechanical damages, including ultrasonics, acoustic emission, eddy current, x-ray, vibrations [2].

The damage detection is essentially concerned with four main functions [2]: (i) the early detection of imperceptible damages, (ii) the localisation, (iii) the severity assessment, and (iv) the residual lifetime evaluation (prognosis). These tasks have been addressed using both model-based and data-driven approaches. Finite Element Model (FEM) [3] and vibration-based damage identification techniques [4] are two commonly used model-based approaches. Generally speaking, model-based methods are feasible in the case of simple geometry structures and invariant environmental parameters. For large-scale and complex structures, the data-driven methods are effective in the extraction of informative multidimensional features from sensor mea-

surements, which are sensitive to damages and robust to environmental changes and noise [5, 6, 7]. With data-driven methods, the damage detection turns into a pattern recognition problem which is often based on statistical and neural approaches [8, 9].

The damage detection techniques use sensor(s) information in order to evaluate the condition and performance of a structure. The evaluation can be either local, i.e. based on the inspection of the defective area, or global, i.e. based on the inspection of the overall structure. Most NDT techniques are local and require scanning when large areas need to be monitored. Ultrasonic inspection and eddy-current testing (ECT) are well-established techniques for local nondestructive crack detection [10, 11]. Vibration-based techniques have potential to evaluate the entire structural performance, and thus they are global [12, 13]. A global technique allows a continuous monitoring of the structure health state, while a local one is usually carried out offline. Graphical processors, thanks to their high computational capability for repetitive tasks, are now used for on-line local NDE [14, 15].

We will study the application of statistical feature extraction methods, namely PCA and LDA, along with the KL divergence to perform primarily tasks (i), (ii) and (iii). Small cracks will be detected, localised and characterised under two frameworks: eddy-current testing and vibration monitoring. ECT is widely used in many industrial fields, like aeronautic, to detect small surface cracks in conductive structure. A 'global' time-domain analysis based on KL divergence and PCA will be applied to ECT signals with low SNR in order to reveal the signature of minor cracks and characterise their severity.

Vibration monitoring (through accelerometers) is widely used in electrical rotating machines in order to monitor mechanical faults in the machines. Bearing faults are particularly of concern in this study. A 'global' frequency-domain analysis based on an adequate preprocessing and PCA/LDA will lead to high-level discriminative features extracted from vibration signals of faulty bearings. The bearing faults will be classified according to their different types and sizes, allowing thus for their diagnosis.

4.2 Motivation and outline

The present studies are motivated by the fact that many industrial and onboard systems structures are threatened by material cracks which start imperceptible but can develop rapidly, leading to catastrophic failures. Minor cracks should be detected and localised as soon as possible for economic and safety considerations. These

structures are generally equipped with sensors and instrumentation which provide the diagnostic information. However, the detection of minor cracks is still a challenging issue: the incipient fault signature can not be detected by visual inspection or simple processing of sensor information. It is usually weak (high SFR), concealed by environmental disturbances, uncertainties, and noise (low FNR).

The ECT technique measures the change in the sensor impedance, and links its increasing amplitude to the presence of a crack into the tested material. However, the sensor impedance is subject to many sources of noise that have the same effect as cracks. These include the surface roughness of the material, the temperature variation, the lift-off [16, 17]. Some denoising techniques have been proposed to improve the SNR [18, 19, 20], but in practice, the measured ECT signals are complex and the origin of the background noise remain unidentified.

Vibration signals in rotating machines are likewise combinations of vibration components from several electromechanical sources. The detection of a particular mechanical fault consists in looking for its spectral signature, i.e. the characteristic frequencies, in the vibration spectrum [21, 22]. These characteristic frequencies should be known in advance. They should be also dominant with respect to the background noise and the numerous frequency components that form the machine vibration spectrum.

The two applications will be investigated separately. Section 3 is devoted to the detection and characterisation of minor cracks based on ECT approach. The divergence is used to reveal the hidden signature of minor cracks in the ECT sensor impedance. The statistical analysis of ECT signals using PCA leads to characterising the cracks dimensions. Section 4 is dedicated to the detection and discrimination of incipient bearing faults based on vibration monitoring approach. An approach for 'global' statistical spectral analysis is proposed in order to detect and identify small bearing faults. The main advantages of this approach is that prior knowledge of the characteristic frequencies is not necessary and besides, few PCA/LDA-based features, specifically only two, will be sufficient to differentiate among different types of bearing faults and different fault severities.

4.3 Application to ECT-based damage detection

4.3.1 Eddy-current testing principle

This method applies to all conductive materials. It is based on Faraday's electromagnetic induction law, stating that a time-varying magnetic induction flux density

induces currents in an electrical conductor [23, 16]. Fig.4.1 illustrates the ECT principle.

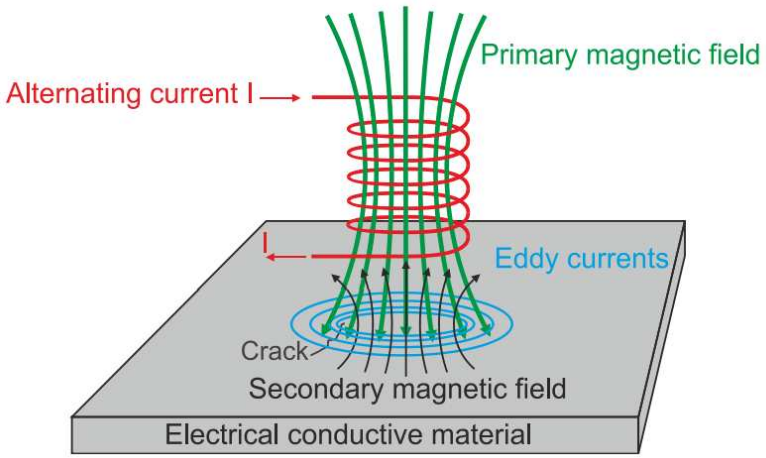

Figure 4.1: ECT principle [16]

The electrical conductor that needs to be tested is placed in a time-varying magnetic field (primary magnetic field) generated by an alternating energized coil. Following the Faraday's law, continuous and circular eddy currents will be generated within the test material. These currents generate a secondary magnetic field that tends to oppose the primary one, thus modifying the electromotive force of the probe coil. The presence of a surface crack in the test material will change the eddy currents flow as shown in Fig.4.2. The secondary magnetic field will consequently change, inducing a change in the electromotive force.

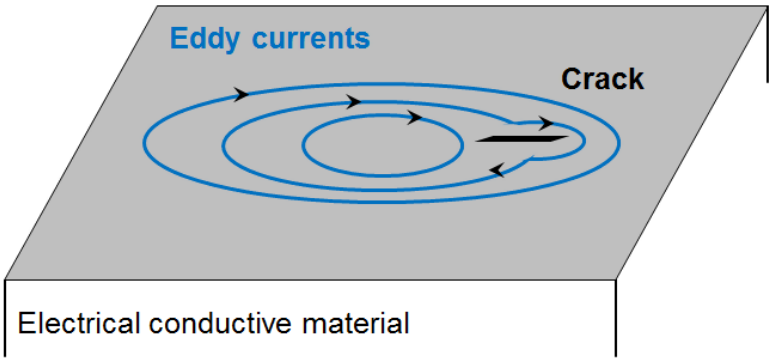

Figure 4.2: Change of the eddy current flow in the presence of a crack

Surface cracks are thus detected by measuring the coil electromotive force variation, or equally the coil impedance variation $\boldsymbol{\Delta Z}$. The coil impedance changes are

the following:

The energized coil taken separately has an impedance
$\boldsymbol{Z_0 = R_0 + jX_0 = R_0 + j2\pi f_{exc} L_0}$,

The set (coil + test material) has an impedance
$\boldsymbol{Z_c = R_c + jX_c = R_c + j2\pi f_{exc} L_c}$

The set (coil + test material + crack) has an impedance
$\boldsymbol{Z_d = Z_c + \Delta Z = R_d + jX_d = R_d + j2\pi f_{exc} L_d}$

where the indexed $\boldsymbol{R}$ and $\boldsymbol{X}$ are the real and imaginary parts of the indexed impedance $\boldsymbol{Z}$, $\boldsymbol{f_{exc}}$ is the excitation frequency, i.e. the frequency of the supplying alternative current, and $\boldsymbol{L}$ is the induction coefficient. So, the ECT inspection consists in measuring the impedance changes $\boldsymbol{\Delta Z}$ that occur in the coil sensor while scanning several points in the tested area. The measurements are taken without any direct physical contact between the sensor and the inspected material, thus permitting a fast surface testing compared to other NDI techniques [24, 25].

The ECT measurements are sensitive to the presence of faults, but likewise to other parameters which should be controlled or taken into account in results interpretation [26, 27, 28]. The distance between the sensor and the inspected material (the lift-off variations), environmental perturbations (temperature variation, electromagnetic fields from external sources, noise, etc.), internal disturbances to the sensor and the inspected material (temperature variation, calibration, internal noise, geometrical changes in the material, etc.) are all factors of the probe coil impedance change. The interpretation of ECT signals may therefore be complex, requiring some skills. The material properties, mainly its electrical conductivity and magnetic permeability, and the excitation frequency driving the probe coil have a direct effect on ECT sensitivity and penetration. The coil type and size also affect the test capability. Therefore, the sensor is usually chosen according to the dedicated application and the required test performance.

4.3.2 Experimental test bed description

The probe-specimen configuration is shown in Fig. 4.3. It was provided by the jet engine manufacturer Snecma [29]. The ECT probe is axisymmetric and constituted of a coil regularly wound around a ferrite core. The diameter and height of the core are 0.8 mm and 4 mm, respectively. Its relative permeability is 1100 and it has a negligible conductivity. The coil has a rectangular cross-section with an external diameter of 1.2 mm and a height of 1.4 mm. It is composed of 110 turns of 40-μm-diameter copper wire. The distance between the bottom of the coil and the bottom

of the ferrite core is 100 μm. The conductive specimen, i.e. the inspected material, is a rectangular shape nickel-base superalloy plate. The thickness of the considered specimen is 3 mm, its conductivity is 0.76 MS/m and its magnetic permeability is almost 1.

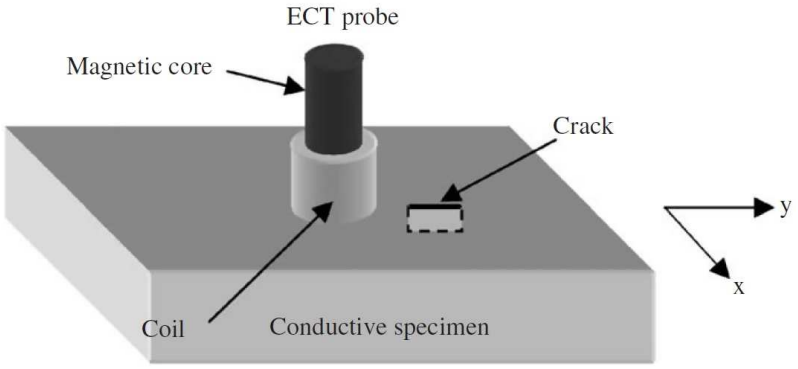

Figure 4.3: Probe-specimen configuration [29]

Several rectangular notches of small sizes were realised in the conductive plate using Electro Discharge-Machining (EDM) in order to simulate thin cracks [30]. Each notch has an opening of 0.1 mm. The considered notches are distributed into 4 lengths: 0.6 mm, 0.4 mm, 0.2 mm, 0.1 mm and 3 depths: 0.4 mm, 0.2 mm, 0.1 mm. The lengths and depths have a 0.02 mm margin. Table 4.2 summarizes the notches dimensions. These notch sizes are representative of the typical size of slight

Table 4.1: Surface, length (l_c) and depth (d_c) of cracks

Area (mm^2)	0.01	0.02	0.04	0.06	0.08	0.12	0.16	0.24
l_c , d_c (mm)	0.1 0.1	0.1 , 0.2 0.2 , 0.1	0.1 , 0.4 0.4 , 0.1 0.2 , 0.2	0.6 , 0.1	0.2 , 0.4 0.4 , 0.2	0.6 , 0.2	0.4 , 0.4	0.6 , 0.4

cracks which have to be detected in many industries.

The probe was moved over the plate with a three axis computer-controlled robot. The probe impedance was measured using an Agilent 4294A Precision Impedance Analyser at the operating frequency of 2 MHz. The experimentation system is shown in Fig.4.4. The probe scanned the different notches in steps of 0.1 mm in both the x and y directions.

Figure 4.4: Experimentation system

The probe coil is driven by a sinusoidal excitation current with a frequency $\boldsymbol{f_{exc}}$ ranging from 0.8 MHz to 6 MHz. Using a single excitation frequency limits the eddy currents penetration depth, and consequently the eddy currents damage detection capability. Low frequency tests allow for a high penetration depth and are suitable for detecting subsurface cracks. High frequencies increase eddy current flow at the surface and are suitable to detect small surface cracks [16]. The low limit frequency (0.8 MHz) is due to the decrease of the probe sensitivity when the frequency decreases, whereas the high limit (6 MHz) is due to the coil resonance frequency.

4.3.3 Data structure and preprocessing

In the ECT data acquisition, the probe is going all over the flawed areas with a spatial step of 0.1 mm in both the x and y directions. Impedance signals are acquired for each crack considered separately. The ECT data, constituted of the real and imaginary impedance parts, are acquired with 7 excitation frequencies $\boldsymbol{f_{exc}} = \{0.8, 1, 1.5, 2, 3, 4, 5, 6\}$ MHz. So for each crack, 7-frequency 2D impedance map centred on the crack is obtained. The map is 40-points long along the direction (y) and 32-points long along the direction (x). The points are 0.1 mm spaced. Fig.4.5 illustrates the maps of imaginary impedance part acquired for serious cracks with a 2 MHz excitation frequency. The impedance variation showing two obvious high-amplitude lobes readily indicates the presence of the cracks with sizes (l_c=0.4mm, d_c=0.4mm) and (l_c=0.4mm, d_c=0.6mm). Nevertheless the detection of smaller cracks is much more tedious. Fig.4.6 shows the impedance variation maps obtained for the smallest cracks, namely the cracks of sizes (l_c=0.1mm, d_c=0.1mm) and (l_c=0.2mm, d_c=0.1mm).

The maps do not show any particular change in the impedance for these minor cracks, making the detection of such faults unobvious. The signature of the fault

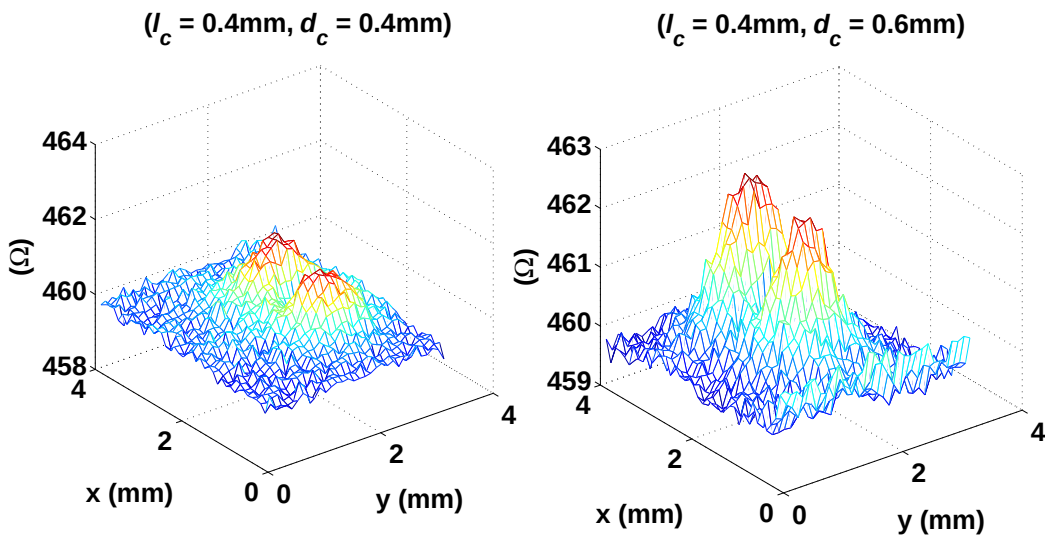

Figure 4.5: ECT map of imaginary impedance for serious cracks

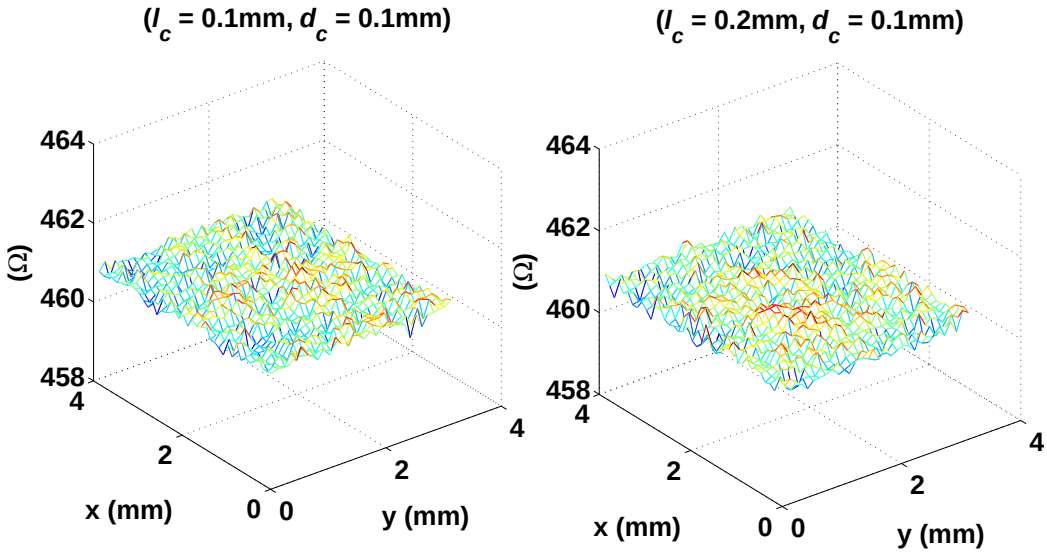

Figure 4.6: ECT map of imaginary impedance for minor cracks

(crack), however, could be masked by the healthy variations. The KL divergence will be used for revealing this signature.

It has been noticed that ECT maps acquired for the same areas at different times

can have slightly different reference impedance which is specified by their respective edges. This is due to changes in environmental conditions to which the eddy currents are known to be responsive. Therefore, all ECT maps are normalised with respect to their reference impedance specified by the edges. Then, the values of each normalised ECT map are arranged into an impedance vector of 1280 (40*32) length. A reference impedance signal is formed with the impedance values of the edges of the normalised maps, since the cracks are at the centre of the maps and the edges are not affected by their presence. This can be shown even with the serious cracks in Fig .4.5.

The ECT technique is usually used for diagnosing specifically large surface structure, since it is a relatively high speed testing technique. Accordingly, we artificially enlarge, to 500*100 points, each normalised ECT map by three steps:

1. Arranging the impedance imaginary part of the ECT map edges into a vector with Additive White Gaussian Noise (AWGN).
2. Concatenating several vectors that have been obtained in such manner.
3. Concatenating the resulting vector with the impedance vector of the ECT map.

The AWGN simulates the potential measurement errors that may occur during scanning large areas. It is chosen with a SNR equal to 20 dB. The SNR could be much smaller (<10 dB) in some cases, for example when inspecting a rough surface. The plate under inspection is nearly smooth and the measurement bench is isolated from the environmental perturbations. For such a case, a SNR of 20 dB seems reasonable. A vector/signal of normalised imaginary impedance, of size $5 * 10^4$, is created as has just been described for each acquired ECT map.

4.3.4 Detection and diagnosis results

4.3.4.1 Crack detection results

For each crack, several signals of the normalised impedance imaginary part are created according to the three previous steps. The detection with the KL divergence $\hat{D}$ computed as Eq.(2.9), consists in (1) estimating the probability distributions of the normalised impedance signals and then (2) comparing them to the probability distribution of the normalised reference signal. Fig.4.7 shows the pdfs estimated for the reference signal and the faulty signal with the crack of sizes (l_c=0.4mm, d_c=0.4mm). The probability distributions are far from normal and using the divergence, the disparities between the two pdfs along their shape will be quantified and amplified. The sample mean, denoted μ, will be calculated in parallel for each

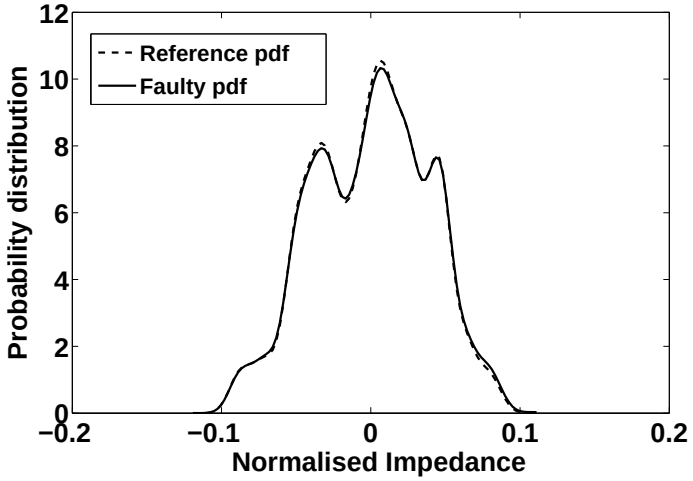

Figure 4.7: Probability distribution of ECT normalised impedance signals

signal. The following procedure is adopted to test the detection capability of $\hat{D}$ and μ:

- 50 realisations (R) of faut-free impedance signals are created. μ is calculated for each fault-free signal. $\hat{D}$ is calculated between the corresponding pdf and the pdf of the reference signal.
- 50 another realisations (R) are created for each crack. μ is calculated for each faulty signal. $\hat{D}$ is calculated between the faulty pdfs and the pdf of the reference signal.
- In order to compare the detection sensitivity of μ and $\hat{D}$, we propose to evaluate, for each criterion, the ratio between the amplitude of the average step due to the fault, and the maximum peak to average value of the criterion. This ratio is given by the following index:

$$Sensitivity = \frac{< Cr >_{R>50} - < Cr >_{R<50}}{max_{Cr}} \tag{4.1}$$

  where Cr is the considered criterion, i.e. μ or $\hat{D}$. The symbol <> denotes the average value, and max_{Cr} is the maximum peak to average value before the fault, see Fig.4.8.

This sensitivity index means that by considering a criterion threshold that provides zero false alarm probability, i.e. a threshold equal to the peak amplitude before the fault as seen in Fig.4.8, we impose zero missed detection probability to be achieved for a *sensitivity* at least equal to 2. Another particular value of this

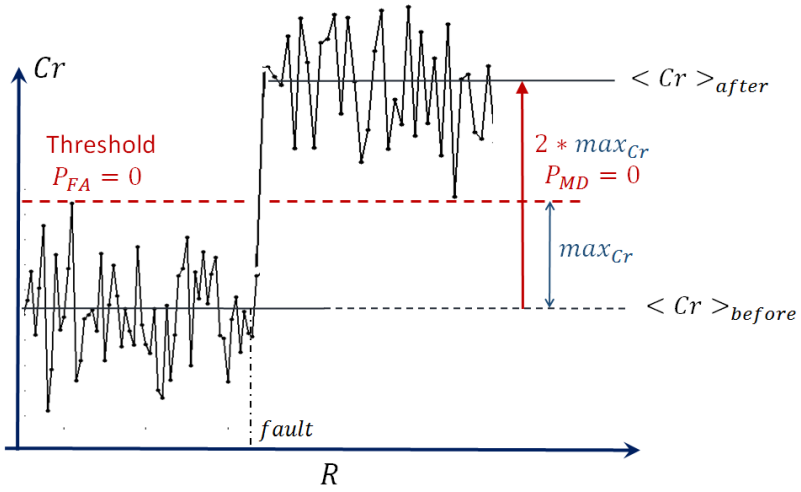

Figure 4.8: Sensibility index

index is 1. We can infer from Fig.4.8 that for a $sensitivity = 1$, P_{MD}=0.5. Then, $0 < PMD < 0.5$ if $1 < sensitivity < 2$.

The detection results are shown especially for the smallest cracks, for which the visual inspection of the ECT maps fails to detect faults (Fig.4.6). By applying consequently the procedure, Fig.4.9 has been obtained. The dashed lines are the thresholds calculated for μ and $\hat{D}$ at the 99% of their maximum values. Obviously, the sample mean fails to detect the cracks with (l_c=0.1mm, d_c=0.1mm) and (l_c=0.2mm, d_c=0.1mm), whereas the divergence is well sensitive to their presence and shows a significant step variation, that is indicated by the arrow. The sensitivity of μ and $\hat{D}$ to the four considered cracks is evaluated and results are given in Table 4.2. The divergence shows better sensitivity than the sample mean to the presence of these small cracks. μ clearly fails to detect the smallest cracks that have (l_c= 0.1mm,

Table 4.2: Sensitivity to minor cracks

Sensitivity / (l_c, d_c) mm	$\frac{< \hat{D} >_{R>50} - < \hat{D} >_{R<50}}{max_{\hat{D}}}$	$\frac{< \mu >_{R>50} - < \mu >_{R<50}}{max_{\mu}}$
0.1, 0.1	5.31	0.26
0.2, 0.1	5.73	0.47
0.1, 0.2	7.37	1.12
0.2, 0.2	23.8	3.64

d_c=0.1mm) and (l_c= 0.2mm, d_c=0.1mm) because the $sensitivity < 1$ which means

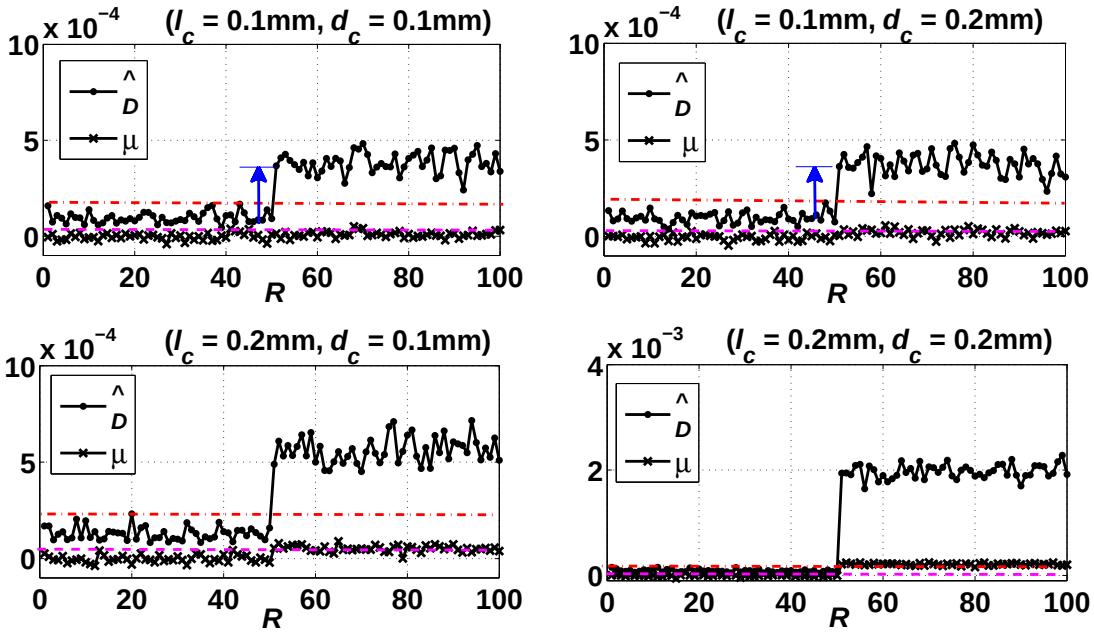

Figure 4.9: Detection of minor cracks with the divergence and the sample mean

the $P_{MD} > 0.5$. μ can only detect the most serious crack that has (l_c= 0.2mm, d_c=0.2mm) successfully and without any missed detection error. This crack is however easily detectable visually on the corresponding ECT map in Fig.4.5. $\hat{D}$ in contrast successfully detects all faults, with zero P_{MD} as high values of sensitivity ($sensitivity > 2$) are achieved.

4.3.4.2 Crack characterisation results

Once a crack is detected, it is important to assess its severity by identifying its size. In the state-of-the-art ECT, the inverse problem of inferring the crack length, depth and size from the probe measurements is resolved using mainly numerical tools that aim at reconstructing the profiles of the cracks, given a mathematical crack model. Here, we attempt to show that given a sufficient training data set, the statistical analysis of the probe measurements lead to assess the cracks severity. The method relies on the fact that the eddy current response to a particular crack depends of the excitation frequency [16]:

- With low excitation frequencies, the eddy currents have an important penetration depth. Thus the eddy current response would be sensitive to the crack surface in z-direction (Fig.4.3).

- With high excitation frequencies, current flow is stronger at the surface and decreases rapidly in z-direction. So the current response would be sensitive mainly to the crack length (in y-direction) as they all have the same width.

The statistical analysis uses information from both low and high excitation frequencies. The procedure consists in the following:

1. Statistical parameters, namely the divergence, the mean, the variance and the maximum value are chosen to statistically describe ECT signals
2. Several ECT signals are generated as described in section 4.3.3 for each crack, and for the low and the high excitation frequency, namely for $\boldsymbol{f_{exc}} = \{1, 6\}$ MHz.
3. The chosen statistical parameters are evaluated for each ECT signal.
4. Two information matrices, X_{lf} and X_{hf}, are formed with the chosen statistical parameters as variables/columns. The rows contain the obtained values for each crack.
5. PCA is applied on each matrix. It results in one latent principal component per matrix summarizing the majority of information ($> 95\%$ of total variance).
6. The projection subspace for cracks characterisation is spanned by the two retained principal components (one for the low frequency and one for the high frequency).
7. Data in X_{lf} and X_{hf} associated with each crack are projected into this subspace.

Before going through the result, it stems relevant to show how the excitation frequency can influence the response of eddy currents to the crack sizes. To this end, the procedure described above is applied to only the three cracks with the same surface 0.04 mm^2 and for all $\boldsymbol{f_{exc}} = \{0.8, 1, 1.5, 2, 3, 4, 5, 6\}$ MHz, instead of the low and the highest one. The first principal component PC_1 summarizes 97% of information. Fig.4.10 depicts PC_1 versus $\boldsymbol{f_{exc}}$. One can notice that the PC_1 values obtained with the low excitation frequencies, namely with 0.8 MHz and 1 MHz, are the same for the three cracks. In other words, PC_1 in the low frequencies is only sensitive to the crack surface which is the same for the three cracks. The discrimination between the 3 cracks according to their length increases with $\boldsymbol{f_{exc}}$. The PC_1 value is the highest for the crack with the most important length.

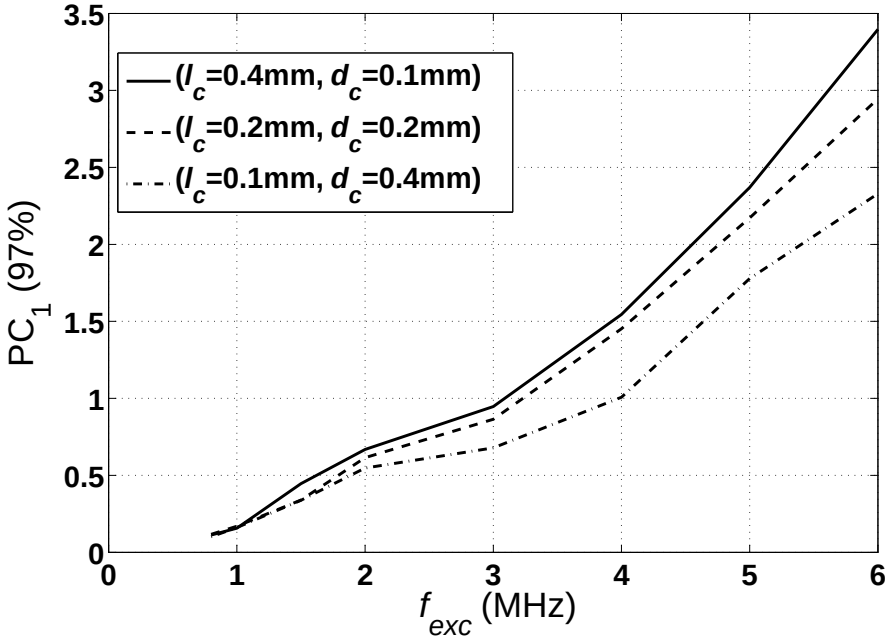

Figure 4.10: The excitation frequency effect

Afterwards, the procedure is applied to 11 cracks at our disposal, for $\boldsymbol{f_{exc}} = \{1, 6\}$ MHz. 40 ECT signals are created for each crack, 20 per excitation frequency, to fill each of the matrices with the evaluated statistical parameters. The different cracks are labeled according to their areas (color label) and dimensions (number label), as described in Table 4.3. The projection of the matrices into the characterisation

Table 4.3: Labeled surface (mm^2) and dimensions (mm) of cracks

Area	0.02 green	0.04 black	0.06 cyan	0.08 red	0.12 purple	0.16 yellow	0.24 blue
l_c , d_c	0.1 , 0.2 (i) 0.2 , 0.1 (ii)	0.1 , 0.4 (iii) 0.4 , 0.1 (iv) 0.2 , 0.2 (v)	0.6 , 0.1 (vi)	0.2 , 0.4 (vii) 0.4 , 0.2 (viii)	0.6 , 0.2 (ix)	0.4 , 0.4 (x)	0.6 , 0.4 (xi)

subspace gives the data scatter shown in Fig.4.11. The principal component PC_1, summarizing the information carried by X_{lf}, is sensitive to the crack surface. Some particular observations are worth to be mentioned:

- Encircled data refer to cracks that have the same surface/area. They have identical PC_1-X_{lf} values.

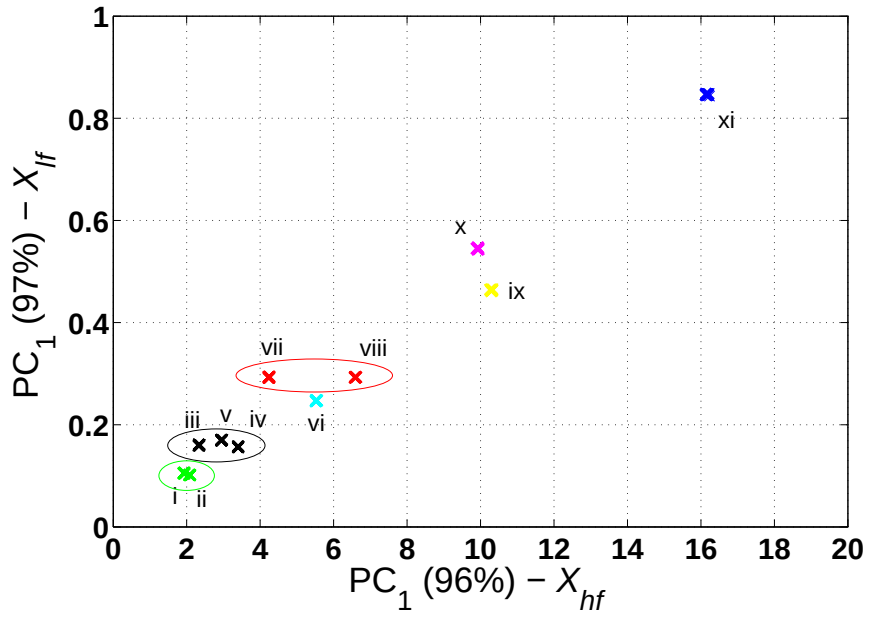

Figure 4.11: Characterisation subspace

- PC_1-X_{lf} increases with the crack surface.
- PC_1-X_{hf} discriminates among cracks that have identical areas, and increases with the crack length.

In real diagnosis, this subspace can be used for the characterisation of new data, by evaluating how the new data are located with respect to classes obtained during the learning process. Therefore, the subspace is supposed to be built for all fault scenarios it is expected to diagnose.

PCA was used to characterise small surface cracks in [30]. Two methods were presented: the first one applies PCA to a matrix containing the maximum of the impedance variation (real and imaginary part), and the second one applies PCA to a matrix that contains the whole cartography. In both cases, the basis matrix of PCA includes the different excitation frequencies and the cracks to be characterised. It was found out that only the first principal component was relevant for the crack charaterisation, as being responsible for 98-99% of the total information. So the projection on the first principal component was correlated to the crack area. The crack dimensions (length and depth) were not characterised. The second principal component highlights some interesting behaviour in the numerical results, but which was not found in the experimentation. In our proposed approach, the obtained principal components are proved to be linked to both the area and length of the

cracks. In future work, it would be interesting to validate, similarly to [30], the present experimental results by conducting the approach on numerical database cartographies.

4.4 Application to vibration-based damage detection

4.4.1 Overview of bearings FDD

The prompt detection and accurate diagnosis of faults in electrical machines has become a main requirement for many industrial and on-board applications. When faults do occur and the machine fails in service, the result could, at best, be the loss of production and income, or, at worst, damage to the industrial process and potentially to the operators and/or the environment. Therefore, the FDD of electrical machines has received increasing attention since the last two decades [31, 32, 33]. The detection of machine faults at an early stage, identifying their locations and the analysis of their causes are essential to ensure the safety, reliability and performance of applications involving electrical machines.

Electrical rotating machines usually operate by means of bearings which are among the most critical components. The quality of the motor system operation is closely related to the performance of bearing assembly. So there are many types of bearings conceived to perform under specific operating conditions defined primarily by the load to support (radial/thrust, small/heavy), the rotational speed (low/high) and the operating cycle length [34]. However bearing faults may account for 42%-50% of all motor failures, as reported in an EPRI (Electric Power Research Institute) publication [35] in 1982 and by Thorson and Dalva [36] in 1999. In fact, bearings, even though properly designed, are sensitive components and their failure is often due to inadequate operating conditions or failures in the maintenance, such as excessive loading, shaft misalignment, wrong mounting, improper lubrication, etc. [37]. In general, one can infer the state of the machine from the state of its bearings.

At very early stage, bearing failure often manifests as small discrete faults on bearing elements, taking various forms such as indentations, spalls, pits or debris. Another type of damage is the case of fatigue, which prematurely results in some surface crack. The typical fault sequence is as follows: spalling on the races, the balls, and finally the cage. Each fault type exhibits characteristic signatures in the time and frequency domains of the machine variables. Among the state-of-the-art, vibration monitoring and motor current analysis are asserted to be the most effective and practical techniques to diagnose bearing faults [32, 38, 39]. Although bearing

vibration signals, which cover displacement, velocity and acceleration signals, are rarely straightforward and may contain vibration components generated by various mechanical and electromagnetic forces, they provide the most salient information for the early detection of bearing faults [40, 41].

Time-domain vibration analysis has the advantage over the frequency-domain analysis, of providing quick, simple and low cost faults detection methods [42, 43]. However, bearing faults as well as other problems including mass imbalance, shaft misalignment, airgap eccentricity and gear failures normally affect the statistical properties of vibrations. Monitoring time-domain fault indicators can only separate healthy from faulty conditions, but has no isolating capability. The investigation of the frequency domain allows to identify the bearing fault type, i.e. its location [44, 45], leading to the cause of failure [34].

Hence, bearing faults, as most of mechanical and electrical faults, have distinguishing spectral features. When spalls or surface crack, appear on a particular bearing element, metal-to-metal contacts occur between rotating bearing elements and the faulty surface. Consequently, the vibration signal will show transient impulses generated almost periodically at a repetition rate characterizing the faulty bearing element [46]. So, different characteristic frequencies can be generated by the bearing depending on which surface is affected by the fault: there is one characteristic fault frequency associated with each of the four bearing elements. Accordingly, there are four types of faults: outer race fault, inner race fault, ball and cage fault. The characteristic frequencies can be computed from the following theoretical expressions [34]:
Ball pass frequency of inner race:

$$f_{inf} = BPFI = \frac{n_b f_r}{2}\left\{1 + \frac{d}{D}\cos\phi\right\} \tag{4.2}$$

Ball pass frequency of outer race:

$$f_{ouf} = BPFO = \frac{n_b f_r}{2}\left\{1 - \frac{d_b}{D_b}\cos\phi\right\} \tag{4.3}$$

Ball spin frequency:

$$f_{bf} = BSF = \frac{f_r D_b}{2d_b}\left\{1 - \left(\frac{d_b}{D_b}\cos\phi\right)^2\right\} \tag{4.4}$$

Fundamental train frequency (cage speed):

$$f_c = FTF = \frac{f_r D_b}{2}\left\{1 - \frac{d_b}{D_b}\cos\phi\right\} \tag{4.5}$$

where f_r is the shaft rotational frequency, n_b is the number of bearing balls, ϕ is the contact angle i.e. the angle of the load from the radial plane, D_b is the pitch diameter and d_b is the ball diameter, see Fig.4.12. For normal speeds the character-

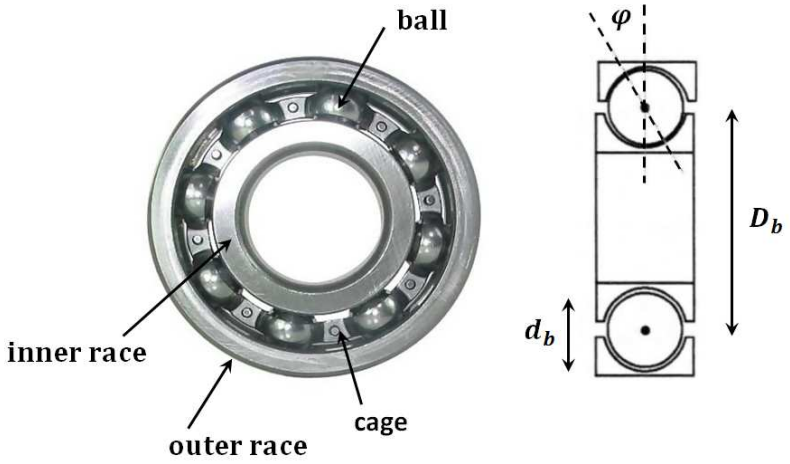

Figure 4.12: Bearings geometrical specifications

istic frequencies lie in the low-frequency range (less than 500 Hz). Also, the train of pulses excites high natural modes in the bearing and the machine, so that an increase in the energy level of the natural frequencies and their sidebands occurs. The resulting time signal typically contains a high frequency component amplitude-modulated at the ball pass frequency [47].

Most of diagnostic methods that use vibration analysis can be roughly classified between classification approaches and vibration measurements. Classification approaches include mainly the use of statistical classifiers, neural networks, support vector machines and fuzzy logic. Detailed discussion of these methods can be found in [48]. Several critical steps are generally involved within faults classification systems. Each one requires considerable attention in order to get successful classification results. The primary step is the feature extraction. It can render the classification algorithm either very complex or quite simple depending on the choice of features. Informative features that have important discriminant power, also called high-level features, are always in need. However, a tradeoff is to be made here with the computational complexity of the pre-processing required to obtain effective features. They involve the use of sophisticated signal analysis techniques such as time-scale analysis [49], time-frequency analysis [50, 51] and denoising techniques [52, 53] which are computationally intensive. Refer to [40] for comparison of the computation complexities among some advanced signal processing methods used for the diagnosis of bearing faults.

Vibration measurement techniques focus on trending the energy level at the ball pass frequencies or in frequency bands related to the ball pass frequencies in order to identify bearing faults. The envelope analysis has been considered as the benchmark method for bearing diagnostics [54, 55, 56, 57]. It consists in demodulating the vibration signal into the region of natural frequencies in order to recover the impulsive signal, indicative of the bearing fault presence. The envelope signal is given as the modulus of the analytic signal obtained using Hilbert Transform [58]. It is then processed into Fourier Transform to identify ball pass frequencies. Lots of works found in the literature propose to combine this method to advanced signal processing techniques in order to further enhance the quality of bearing vibrations. Some useful techniques are reviewed in [46]. However, these works are almost all based on the assumption that the characteristic fault frequencies are known or estimated in advance, in order to elaborate spectral features and fault indicators. The effectiveness of those techniques depends on the accuracy of the ball pass frequencies estimation. These frequencies are either given by the manufacturer for a new bearing, or theoretically estimated using the geometric dimensions of the bearing and the rotational speed as given in (4.2)-(4.5).

4.4.2 Problem statement and contribution

The diagnosis of bearing faults usually requires the prior knowledge or estimation of the characteristic fault frequencies. Unfortunately, the geometric dimensions of bearings from which the characteristic fault frequencies are calculated, can be in practice unknown, for pre-installed bearings for example. Besides, the difficulty arises when the actual characteristic frequencies differ from the theoretical estimated ones due to various factors such as ageing, unexpected rotor slip and excessive friction, change in the ball contact angle and fluctuations of rotor speed [59]. In addition, in real situations and under some conditions of load distribution and rotor speed, the fundamental characteristic frequencies may not be present into the signal spectrum although the fault is quite serious [60]. Instead, several multiple harmonics of these frequencies and sidebands linked to modulations by the rotational frequencies (shaft speed and cage speed) are always pronounced [61]. The global spectral signature of a bearing fault with a characteristic frequency f_{fa} is:

$$nf_{fa} \pm \eta f_c \pm \kappa f_r \tag{4.6}$$

where n, η and κ are integers often $\in \{0, \pm 1, \pm 2, \pm 3\}$, f_c is the rotational frequency of the cage and f_r is the rotational frequency of the machine. However, the occurrence of a particular harmonic or a sideband is a random process and its level depends on the fault itself and likewise other unpredictable causes [60, 34] such as

the operating load condition and the presence of other machine problems like load imbalance, shaft misalignment, looseness, etc.

Although the ability to detect the characteristic fault frequencies is necessary, this approach neglects a large part of the global signature lying throughout the signal spectrum. Some exceptions to these works can still be found. For example, in [62, 63] some discriminating frequencies are extracted, and then they are used to recognize bearing faults. Complex classifiers are required to make bearing faults classification on the basis of large amounts of vibration data related to each fault condition. Trajin et al. define in [40] a spectral energy indicator for bearing faults based on the energies extracted from the spectra of vibration signals and contained into specific frequency ranges related to the theoretical characteristic frequencies. The ranges include basically multiple harmonics of ball pass frequencies and modulations linked to the mechanical speed and cage frequencies. The detector however requires accurate estimation of the ball pass frequencies for a given rotational frequency. Moreover, it is not able to discriminate between inner race faults and outer race faults because the associated frequency ranges exhibit overlaps.

So, it can be argued that, in practice, characterising each fault type with a single frequency component to make the diagnosis can be misleading. On one hand the occurrence of the fault characteristic frequency is not guaranteed in real situations, and on the other hand there is a high probability for this frequency to be concealed by other stronger frequencies. Furthermore, the characteristic fault frequencies can be known or accurately estimated only for a new perfect bearing. Moreover, a large part of the global spectral signature produced by the fault is neglected by trending only the energy at the characteristic fault frequencies. The global signature is theoretically formed with several multiple harmonics of the fault frequencies and sidebands related to their modulations by the rotational frequencies. Nevertheless, these frequencies are actually unpredictable, since the occurrence of a particular harmonic or a sideband is actually a random process.

For the above reasons, we propose a fault classification scheme, which overcomes the usual prior knowledge of the characteristic frequencies. Attention will be paid to the salient information revealed by features related to the global signature that a natural bearing fault may exhibit into the vibration spectrum: spectral lines at ball pass frequencies, their multiple harmonics and their modulations by other frequencies. These spectral lines can not be extracted from the direct spectrum without the knowledge of their exact position, because the vibration spectrum is usually very complex, as being a continuous row of many frequency components from multiple sources. However, the spectrum of the envelope vibration signal, that can be

recovered using the envelope demodulation technique, is exempt from most of the useless frequencies. The frequency components related to bearing damage are quite prominent in the envelope spectrum and most of them could be obtained by picking out spectral lines with the highest amplitudes.

So we here deal with the discrimination between the conditions of bearing faults based on a global spectral analysis. This global spectral analysis allows to obtain spectral features with significant discriminatory power. These features are extracted from the envelope spectra of vibration signals without prior knowledge of the bearings specific parameters and the characteristic frequencies. The extracted spectral features form the global spectral signature produced by the bearing faults. The global analysis method combines performances of the envelope analysis, the sliding FFT and the principal component analysis to establish the faults classification space. No complex statistical classifiers are involved in the classification.

Usually, the signature of faults in the bearing balls is very weak and hard to be detected and identified. Moreover, the diagnosis of faults in the bearing balls is less broached in the literature compared to the other bearing faults types. This is due to the fact that introducing an artificial defect into the races is easier than into the balls, and besides, the ball faults signature is always very weak because of the double rotating motion of the balls. Therefore, this work proposes to enhance the diagnosis of ball faults, using the Linear Discriminant Analysis as part of the global spectral analysis method. The method will be applied to experimental vibration data acquired from bearings containing different types of faults with different small sizes. The Bhattacharya distance and the classification error rates will be evaluated in order to confirm the efficiency of the obtained results.

4.4.3 Method description

4.4.3.1 The global spectral analysis description

The feature vector, of dimension m, will be formed with the absolute amplitude of spectral lines at the ball pass frequencies, their multiple harmonics and the sidebands related to their modulations by other frequencies including mainly the rotational frequency and the cage frequency. These spectral lines are automatically extracted from the spectrum of the envelope vibration signal corresponding to each fault condition, by selecting the dominant frequency components with respect to a specified threshold. The formed vector contributes, using the Fast Fourier Transform (FFT) technique, to construct an amplitude matrix X carrying information from the different fault conditions of interest. To fill the matrix X, there is no need to have a huge amount of data corresponding to each fault condition. Besides, each

envelope signal is decomposed into several data segments using a sliding window of samples. The FFT is computed on the different segments to get the amplitudes of retained spectral features. Afterwards, Principal component Analysis (PCA) is applied on the spectral matrix of information to get a lower dimensional principal subspace highlighting the differences and the similarities contained into X. This subspace represents the output space for classification and diagnosis.

Fig.4.13 summarizes the global approach. It comprises of four main steps:

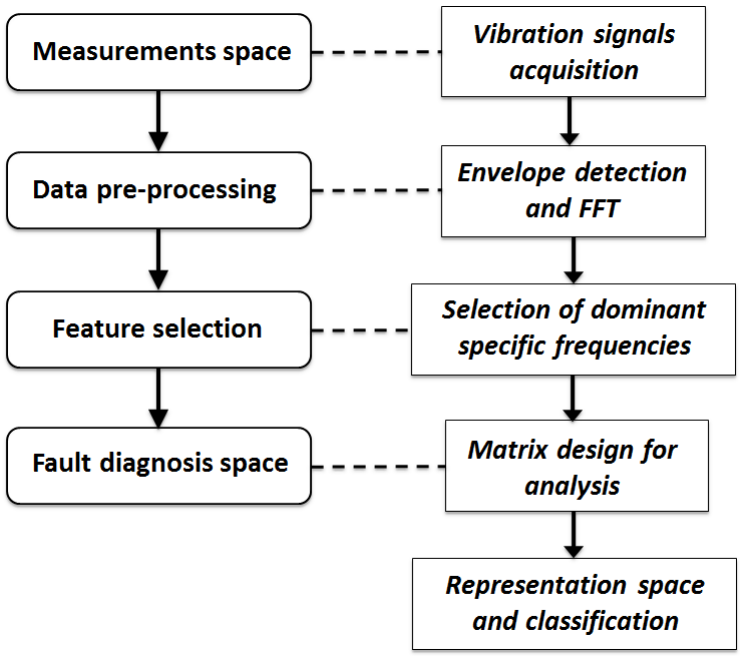

Figure 4.13: The global spectral analysis description

- Step 1: Measurements space. It contains vibration signals corresponding to all conditions of interest: the healthy and all faulty conditions of the bearings. Signals acquired from several operating points of the machine can be included in order to get a general behaviour space for diagnosis.

- Step 2: Data pre-processing. It consists mainly in the envelope analysis. The vibration signal corresponding to each fault condition is first band-pass filtered around the region of natural frequencies. The spectral kurtosis, the kurtogram or its recent improvements [64, 65] can be used to select the bearing resonant frequency bands. However, the difference in dB-spectra compared

to the kurtogram has been shown effective enough to identify the frequency region of interest [66, 46]. It represents the difference between the dB-spectra of a healthy vibration signal and a faulty one, and it just requires historical data to be available. The envelope signal is then given as the modulus of the analytic signal obtained using Hilbert Transform. FFT is applied to each envelope signal in order to extract the spectral features from its spectrum.

- Step 3: Spectral features selection. The line spectrum of each envelope signal is scanned to pick out only the frequencies of highest amplitude. The set of dominant frequencies associated with each fault condition is supposed to include the specific fault frequencies, i.e. the ball pass frequency, several harmonics and sidebands related to their modulations by the rotational frequencies. The rotational frequencies and their multiple harmonics can be selected during the extraction. However, all the common frequencies are withdrawn from the total set of features in order to keep only the distinguishing components. The scanned frequency range can be reduced, since for normal speeds the characteristic fault frequencies are usually less than 500 Hz.

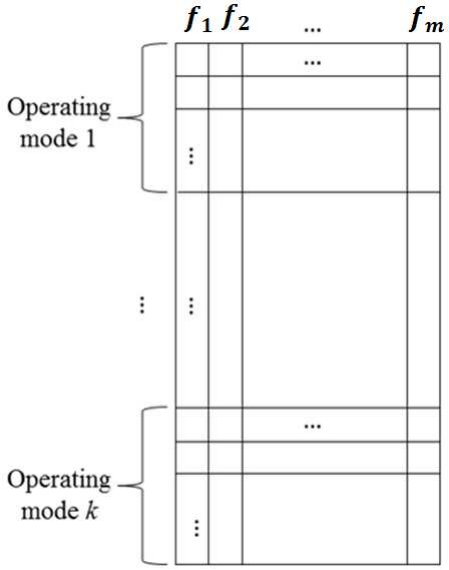

Figure 4.14: Matrix design for analysis

- Step 4: Fault diagnosis space. The X matrix is formed as shown in Fig.4.14 with m retained frequencies $\boldsymbol{f_i}$, $\boldsymbol{i} = \{1, ..., m\}$, considered as variables or fea-

tures. It contains several (k in Fig.4.14) sub-matrices each referring to an operating point of the machine, *i.e.* a particular fault type with a particular severity level. Each envelope signal is decomposed into several data segments using a sliding window of samples in order to construct the associated sub-matrix in X. At each position of the window, the FFT is computed on the corresponding data segment and the absolute amplitudes of the retained variables are acquired and arranged into X as a row sample. The size of the window is chosen with reference to a trade-off between the frequency resolution and the number of samples in the sub-matrices. PCA is then applied to the covariance matrix of X to get the output PCA space for classification.

4.4.3.2 Discrimination of faults in the bearing balls using LDA

PCA and LDA are linear transformations that can be used to perform dimensionality reduction. However, theoretically, LDA leads to better data classification than PCA: PCA performs dimensionality reduction while preserving as much of the total variance in the high dimensional space as possible whereas LDA performs dimensionality reduction while preserving as much of the class discriminatory information as possible. This analysis technique considers maximising the following objective, also called Fisher criterion [67]:

$$J(u) = \frac{u^T S_B u}{u^T S_w u} \tag{4.7}$$

where S_B is the between classes scatter matrix and S_w is the within classes scatter matrix. If the number of classes to be discriminated is c, the LDA projects the space of the original variables onto a $(c-1)$ - dimensional space which axes u are obtained by maximising (4.7). Intuitively, this operation is a compromise between maximising the distance between the projected centers of classes, and the minimisation of their variances, thereby facilitating classification. By using Lagrange multipliers, the solution to this problem yields u as the eigenvector of $S_w^{-1} S_B$ associated with the largest eigenvalue. The eigenvectors associated with the largest eigenvalues define discriminating axes which span the LDA space for classification. For any c - class problem we would always have $c-1$ non-zero eigenvalues.

In the global spectral analysis method, LDA will substitute PCA in order to improve the discrimination among different sizes of ball faults [68]. So, LDA will be applied to the data set excluding data of races's faults. As shown in Fig.4.15, only the samples corresponding to the healthy and the ball faults (BF) conditions (including all operating points) are considered to build the LDA transformed space. The original space is still formed with the specific frequencies $\boldsymbol{f_i}$ for $\boldsymbol{i} = \{1, ..., m\}$,

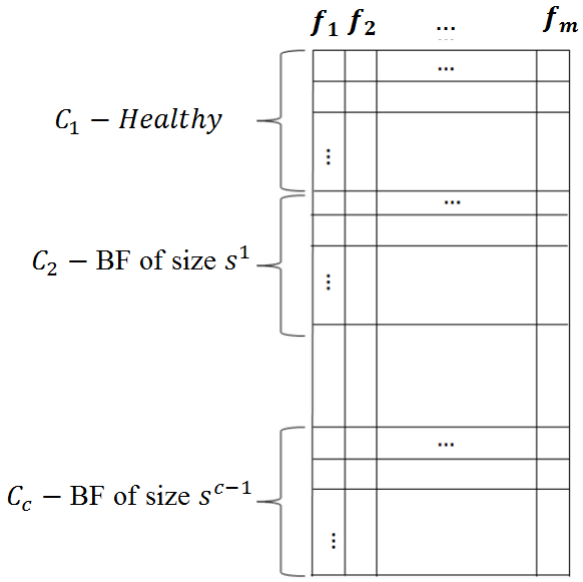

Figure 4.15: Data set design for LDA

that were extracted according to the diagram given in Fig.4.13.

Mathematically formulating the transformation, consider N samples $\{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ of dimension m where N_i of the samples belongs to class C_i, $i = 1, 2, ..., c$. Let μ_i be the mean of class C_i and μ be the mean of entire data given by:

$$\mu_i = \frac{1}{N_i} \sum_{\boldsymbol{x_n} \in C_i} \boldsymbol{x_n}, \qquad \mu = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x_n}. \tag{4.8}$$

The between classes scatter matrix can be estimated as:

$$S_B = \sum_{i=1}^{c} (\mu_i - \mu)^T (\mu_i - \mu) \tag{4.9}$$

The within classes scatter matrix is estimated as:

$$S_w = \sum_{i=1}^{c} S_i \quad \text{where} \quad S_i = \sum_{\boldsymbol{x_n} \in C_i} (x_n - \mu_i)^T (\boldsymbol{x_n} - \mu_i) \tag{4.10}$$

The eigenvector decomposition of $S_w^{-1} S_B$ yields the transformed LDA space as spanned by the eigenvectors corresponding to non-zero eigenvalues. Projecting the original data set $\{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ onto this space allows the classes separability.

4.4.4 Experimental data

4.4.4.1 Experimental test bed description

The proposed methodology has been proved on experimental data. As an example, we present here the results obtained using the Case Western Reserve University (CWRU, Bearing Data Center) vibration data [69]. According to the description

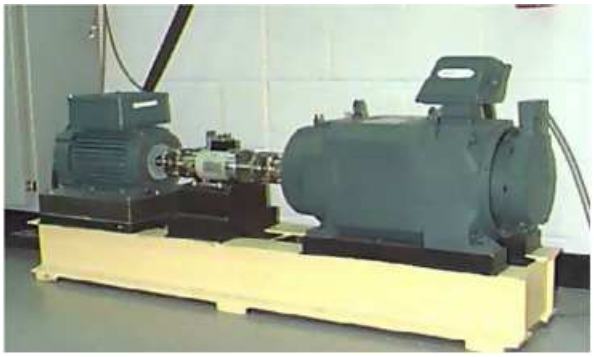

Figure 4.16: Test-bench of Case Western Reserve University

given by the provider of the test data, the test stand, shown in Fig.4.16, consists of a 3hp motor (left), a torque transducer/encoder (centre), and a dynamometer (right). The test bearings, including drive end and fan end bearings, support the motor shaft. Vibration data were collected using 3 accelerometers, which were attached to the housing with magnetic bases at 3, 6 and 12 o'clock positions. 6 o'clock position is located in the load zone, 3 o'clock is orthogonal to load zone, and 12 o'clock is in the bearing clearance zone. Results are shown here for vibration signals from the orthogonal accelerometer.

Single point faults are introduced separately at the inner raceway, the balls and the outer raceway using electro-discharge machining (EDM). SKF 6205 series deep groove ball bearings are used in the experiments. The geometrical specifications of the bearing are: ball diameter d_b= 7.94 mm; pitch diameter D_b= 39.04 mm; number of balls n_b= 9; and contact angle ϕ= 0.

Vibration signals are acquired, at a sampling frequency of 12 kHz during 10 sec, using an accelerometer mounted on the housing of an induction motor system coupled to a load that determines the motor speed (step 1). A 2 Hp Reliance Electric motor is used, and experiments are repeated for the motor being unloaded (0%), 50% loaded and full loaded (100%).

Three types of faults, *i.e.* outer race fault ORF, inner race fault IRF and ball fault BF, and two sizes (180 μm and 530 μm) for each fault are considered, so that

the dataset can be divided into 7 classes including one class for healthy bearings and one class for each fault condition. For each class, there are three 10-second vibration signals each associated with an operating point. For the LDA application, three classes are considered including the healthy class and the two classes for two sizes of faults in bearing balls. The faults created can be considered incipient: the smallest hole (0.18 mm) represents 1/44 of the ball diameter (d_b= 7.94 mm) and the more serious one (0.53 mm) is of 1/15.

4.4.4.2 Time-domain detection

Even a perfect bearing produces vibrations when loaded, due to the rotation of its elements [60]. They are called varying compliance vibrations. The healthy vibration signal however can be assumed as stationary random process and simulated as Gaussian noise. When metal-to-metal contacts occur between rotating elements and the faulty surface, the vibration signal shows transient impulses. Variance analysis (Root-Mean Square (RMS) level), crest factor which is the ratio of peak-value to RMS, Peak-to-Peak analysis, kurtosis and skewness analysis, are commonly used to detect the non-Gaussian components of vibration signals [70, 71]. They are either used to detect faults on the basis of a threshold that corresponds to Gaussian distribution, or fed to a pattern recognition-based classification system in order to make faults diagnosis.

The KL divergence can likewise be used to detect the deviation of the vibrations distribution from its Gaussian reference. Healthy vibrations, from which the reference distribution can be calculated, must be available. The probability distributions should share the same support, for being able to compute the divergence. This is plausible in case of faults that have weak signature, which introduces distortions along the distribution shape without changing its form or shifting it. Otherwise, the KL information has to be computed instead of the divergence. Fig.4.17 shows the vibration signals at the nominal load acquired from a healthy bearing and a bearing with a 0.007-inch ball fault. The ball fault produces high amplitude vibrations. The IR and OR faults produce vibrations with greater amplitude, due to the shorter transmission path between the shock position and the transducer. Fig.4.18 displays the probability density functions estimated for the healthy, 0.007-inch BF and 0.007-inch IRF vibrations acquired at the nominal load level. Scalars, such as the variance or the peak-to-peak value, are able to detect such faults. KL information can likewise be used, although it involves a higher computational cost. The healthy pdf (continuous pdf line) is normal and it is set as reference for computing the information. Each vibration signal is decomposed into 20 segments, and the KL information is computed on the pdf estimated for each segment. Fig.4.19 shows the results obtained for the 3 BF and 3 healthy vibration signals at our disposal. The KLI computed on successive vibration data segments gives near-zero values

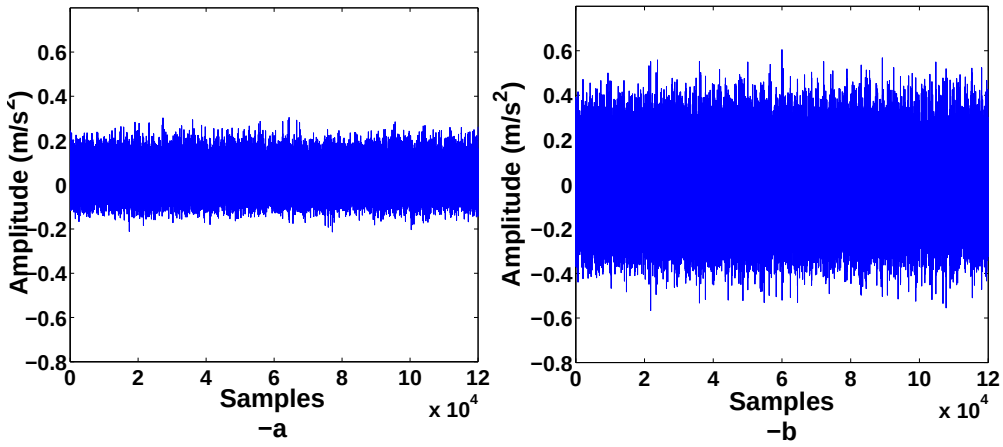

Figure 4.17: Vibration signals at the nominal load -a: Healthy -b: 0.007-inch ball fault

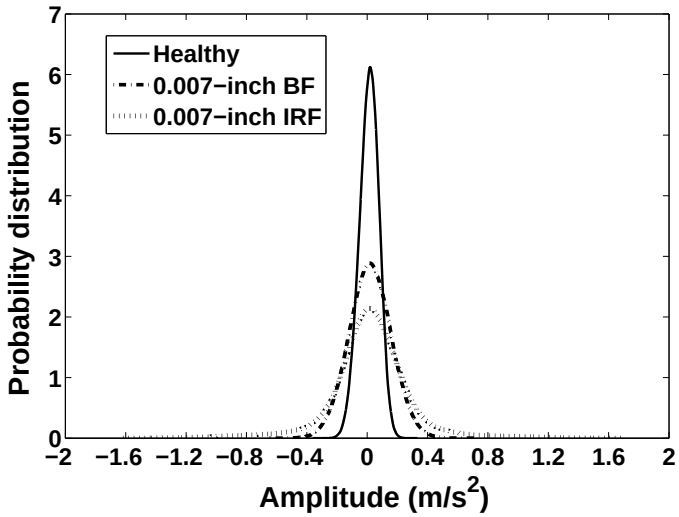

Figure 4.18: pdf

for healthy vibration signals, and relatively large values for faulty vibration signals. However, as it has been mentioned, the difficulty is rather to localise than to detect the bearing fault. In the sequel, the vibration signals will be processed according to the global approach we propose in order to make the diagnosis.

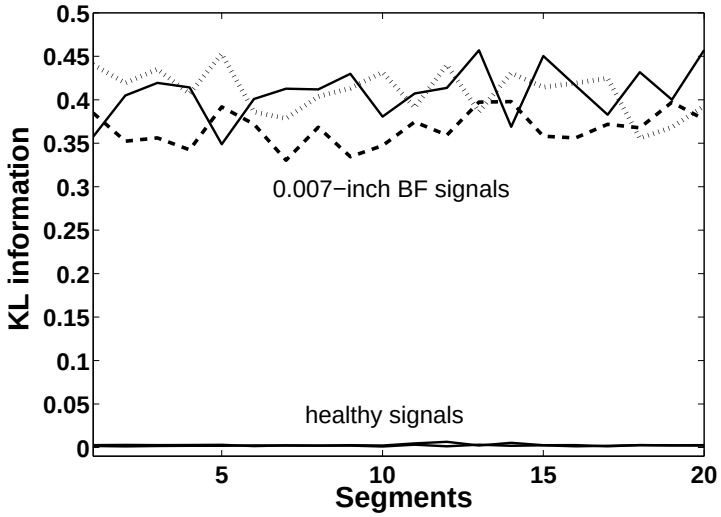

Figure 4.19: KL information

4.4.4.3 Data preprocessing

The pre-processing (step 2) consists mainly in the envelope analysis. This technique has been considered as the benchmark method for bearing diagnosis because it aims at recovering the impulsive signal indicative of the bearing fault presence. The fault characteristic frequencies, if they are already known, should show up into the envelope spectra. The ball fault characteristic frequency, so called ball spin frequency (BSF), is theoretically given by:

$$BSF = \frac{f_r D_b}{2d_b}\left\{1 - \left(\frac{d_b}{D_b} cos\phi\right)^2\right\} \tag{4.11}$$

where f_r is the shaft speed. Fig.4.20 depicts the envelope spectra corresponding to ball faults, the motor being 100% loaded and rotating at 1750 rpm. The BSF calculated using the specifications of the bearings and the considered rotational speed gives 68.88 Hz. Conventionally, the BSF or the even harmonics ($2BSF$) should be dominant in the envelope spectra, which is not exactly satisfied in this case as shown in Fig.4.20. For a fault size of $180\mu m$, some frequency components close to BSF (69.95 Hz) and $2BSF$ (139.8 Hz) show up instead of the fault frequencies and they are not even dominant. The fault frequencies are barely noticed in the spectrum associated to 530 μm fault size.

Therefore, the diagnosis of faults in the bearing balls can not simply rely on the occurrence of the fault characteristic frequencies. However, the application of

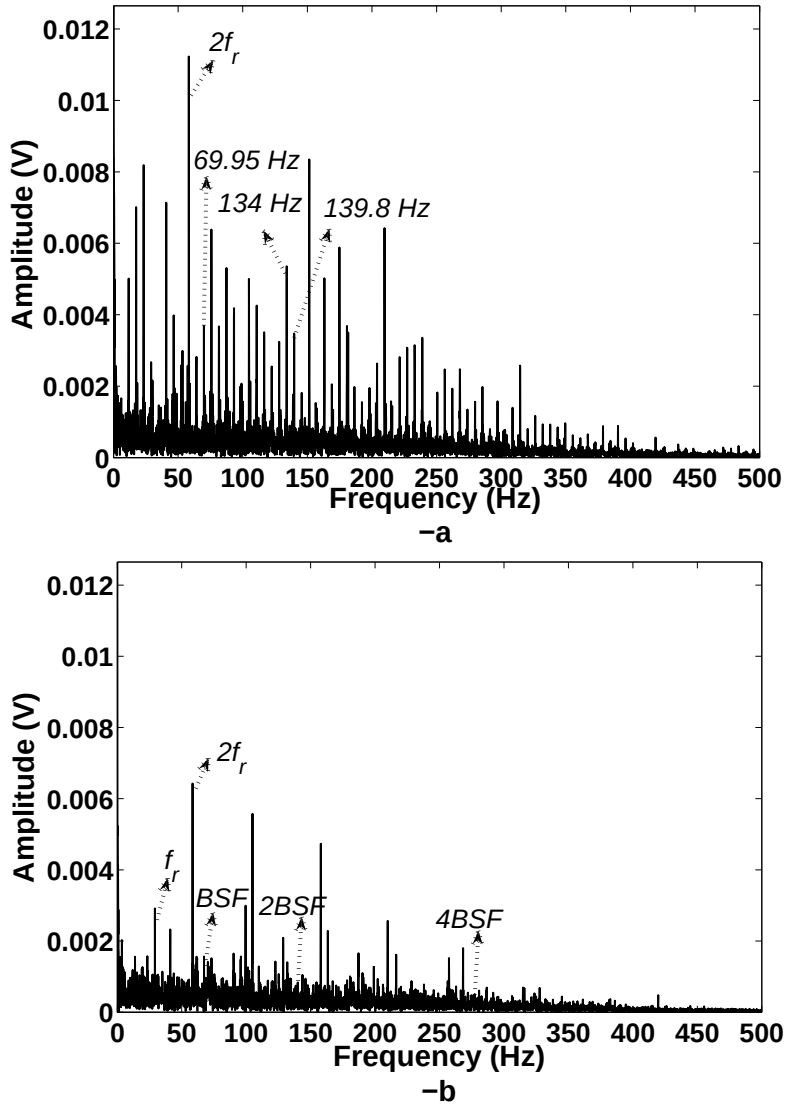

Figure 4.20: Envelope spectra corresponding to BF of size - a : 180μm,- b : 530 μm

the proposed method along with the Linear Discriminant Analysis will be able to discriminate the faultless against the faulty bearings and to distinguish between the two fault sizes without even the estimation of the fault frequencies.

4.4.5 Results with PCA

One of the advantages of the proposed method is that it can easily include multiple operating points. The data available here concern the motor being unloaded (0%), 50% loaded and full loaded (100%). 45 dominant frequencies are extracted from the

available data (step 2), 15 frequencies per operating point. The window of samples slides along each envelope signal with 30% overlap between two successive positions, so that 140 data segments are obtained from each signal. The data segment represents 5 motor rotation cycles (step 3). For each data segment, the FFT is computed and the absolute amplitudes of the retained 45 features constitute a row sample into the X matrix (steps 4). As a result, the spectral matrix X consists of 7 sub-matrices each containing 140*3 samples which represent a fault condition, i.e. a fault type and size.

PCA reduces the 45 - dimensional space into a 2 - dimensional principal subspace spanned by 2 principal components which preserve 89% of the total variance. Projecting the data into the PCA space gives the classification displayed in Fig.4.21. Seven classes are obtained and separated with reference to the fault type and size

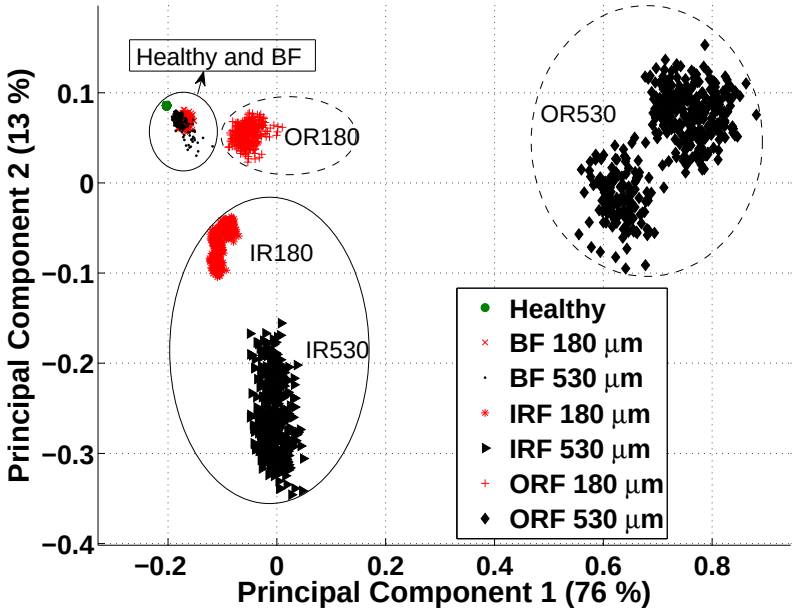

Figure 4.21: Discrimination of bearing faults

whatever the operating point. The classes related to ORF, IRF and BF are well separated and discerned from the healthy class. The classification differentiates between the two sizes of the faults in the races. However, the classes related to ball faults overlap and the discrimination between the two fault sizes is poor.

Fig.4.22 represents a zoom into the PCA space on healthy and BF data. The 2-D probability density function (pdf) for each of the BF classes is estimated and the contours of the obtained pdfs are displayed in Fig.4.23. The 2-D pdfs exhibit an

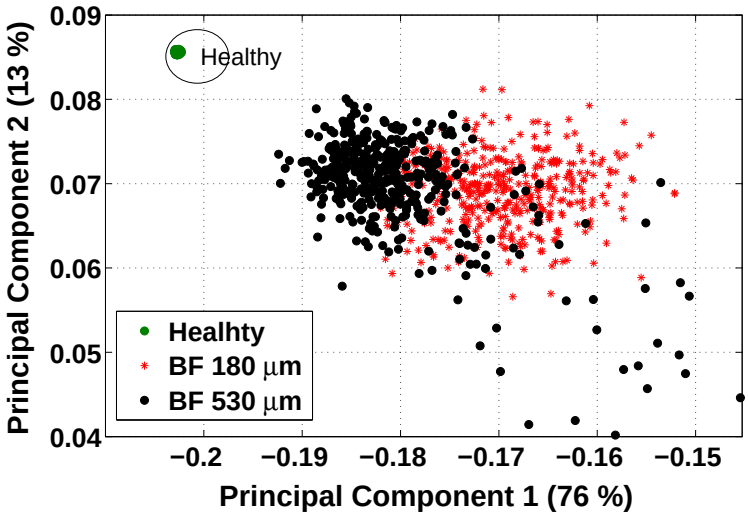

Figure 4.22: Discrimination of ball faults into the PCA space

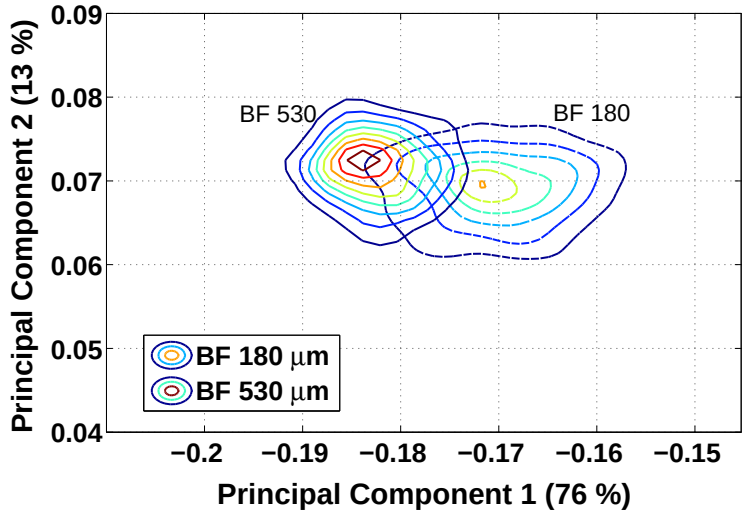

Figure 4.23: Contour of pdfs estimated for the BF classes into the PCA space

important region of intersection which implicitly induces a high rate of classification error. Table 4.4 displays the leave-one-out cross validation rates associated to the classification of ball faults into the PCA space using a linear discriminant boundary between the two classes. The linear boundary is applied to the BF classes projected into the PCA space, where there is a loss of 11% of information. Surely, a quadratic

Table 4.4: Confusion matrix (PCA space)

	$C1$-Healthy	$C2$-BF180	$C3$-BF530
$C1$-Healthy	420 (100%)	0	0
$C2$-BF180	0	326 (77.6%)	94 (22.4%)
$C3$-BF530	0	58 (13.8 %)	362 (86.2%)

or any kernel function can likewise be used to separate between the PCA classes. However, a linear boundary is chosen to be compared to the result that will be obtained with the LDA.

4.4.6 Results with LDA

The data set designed as described in Fig.4.15, contains here 3 classes: C_1 that gathers data from healthy bearings, C_2 and C_3 that represent BF with 2 sizes, 180 μm and 530 μm respectively. Data from the three considered operating points are included into each class. The information into the classes is formed with the same 45 spectral features already used to perform PCA. Each class contains 140*3=420 samples representing the absolute amplitudes of the specific frequencies, for the three conditions of motor load.

The LDA reduces the original 45 - dimensional space into a 2 - dimensional space which, contrary to the PCA space, does not induce any loss of information. Two discriminant axes, also called Fisher axes, are obtained. Fig. 4.24 shows the projection of the data set into the LDA space, and their estimated 2-D pdfs. The eigenvalue associated with each of the discriminant axis indicates its discriminatory power; the first discriminant axis is responsible for 95 % of discrimination between the classes, and the second one for 5%. Compared to those obtained previously, the contours of the pdfs depicted in Fig.4.25 are better separated and show a narrower intersection. Table 4.5 displays the leave-one-out cross validation rates estimated into the LDA space. The cross validation errors decrease from 22.4% to 9% and

Table 4.5: Confusion matrix (LDA space) [72]

	C1-Healthy	C2-BF180	C3-BF530
C1-Healthy	420 (100%)	0	0
C2-BF180	0	382 (91%)	38 (9%)
C3-BF530	0	23 (5.5%)	397 (94.5%)

from 13.8% to 5.5%, thus confirming the contribution of the discriminant variables

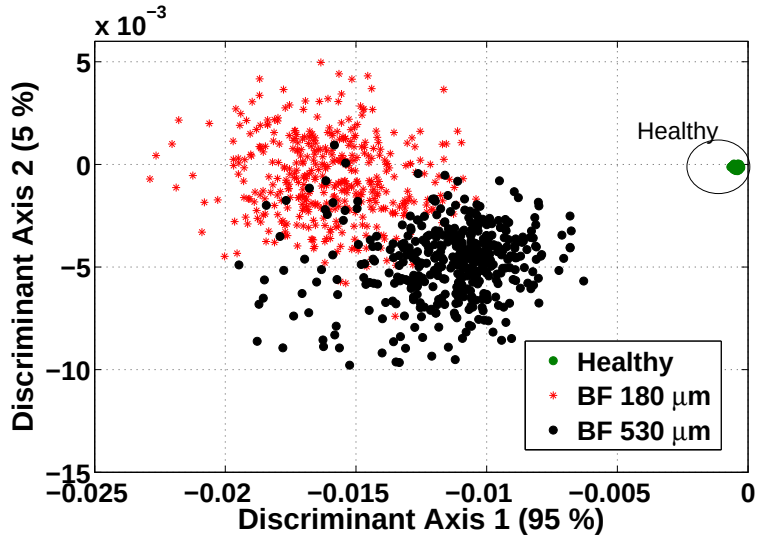

Figure 4.24: Discrimination of ball faults into the LDA space

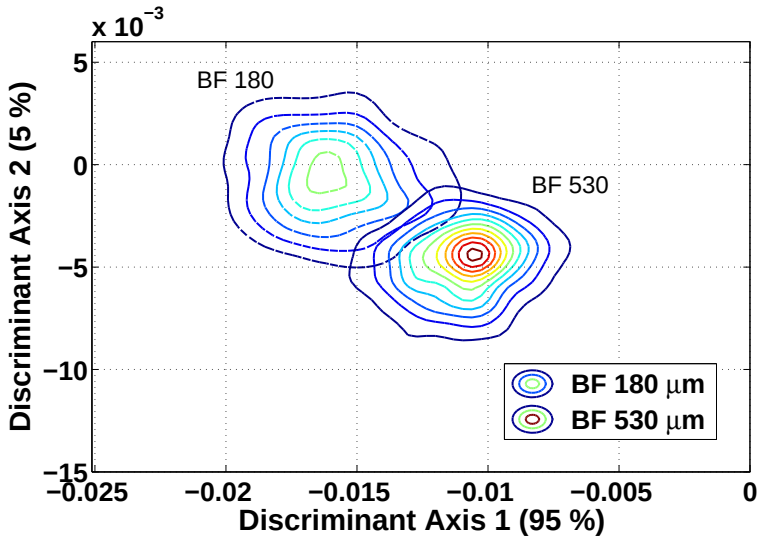

Figure 4.25: Contour of pdfs estimated for the BF classes into the LDA space

to the classification of ball faults.

Beside the cross validation errors, a measure of class separability can be used to quantify the improvement achieved in the discrimination between BF sizes. A well-known class separability measure is the Bhattacharyya distance. It has been

used for pattern recognition and feature selection [73], and for the estimation of the classification error [74]. For two normally distributed classes, the Bhattacharyya distance is defined as follows:

$$B = \frac{1}{8}(\mu_2 - \mu_1)^T \left[\frac{S_1 + S_2}{2}\right]^{-1} (\mu_2 - \mu_1) + \frac{1}{2} ln \frac{\mid (S_1 + S_2)/2 \mid}{\mid S_1 \mid^{\frac{1}{2}} \mid S_2 \mid^{\frac{1}{2}}} \tag{4.12}$$

where μ_i and S_i are the mean vector and covariance matrix for class C_i. Normally, this distance evaluates the separation of the classes and the higher its value is, the more the classes are separated.

Eq. (4.12) is used to quantify the separation between the BF classes after their projection into the PCA and LDA spaces. This assumes that the projected classes are normally distributed, which could be inaccurate, since the distributions are not perfectly normal as shown in the figures above. However, it gives a simple and straightforward tool to make conclusion. In addition, each class contains here only 420 samples. For a larger size, the distributions would be closer to normal.

The Bhattacharyya distance, calculated between the two BF classes, gives:

- into the PCA space: $B = 0.39$
- into the LDA space: $B = 1.1$

The Bhattacharyya distance between the classes in the LDA space is 3 times greater than the value obtained in the PCA space. Therefore, these values confirm that the discrimination of ball faults according to the fault size has been improved using LDA. The LDA extracts from the original spectral features a few discriminating features more adequate than the principal components to make the discrimination between different sizes of faults in the bearing balls.

Many works that propose fault diagnosis approaches use the same experimental data as used in this work to validate their methods. The most recent work uses a combination of singular spectrum analysis (SSA) and artificial neural network (ANN) to classify bearing faults according to their types and sizes [75]. The authors therein compare their method with the recent published works. SSA is applied to vibration signals from each bearing condition. Features that are fed to a feed-forward back propagation neural network, are empirically chosen as being the singular values that distinguish the best the different bearing conditions. The classification accuracy varies from 95% to 100%. The presence of noise contributes to more accurate classification, which seems somehow a strange finding and still not justified in the study. Neural network model development, beside being empirical, is a computationally intensive procedure that requires a great computational time. In our work, the frequency features that are selected as the dominant components

into the envelope vibration spectra are discriminative enough so to avoid resorting to complex classifiers. Equivalent classification accuracy has been obtained while using simple multivariate statistical methods that helps reducing the data dimensionality.

4.5 Conclusion

In engineering systems, the incipient material damages have a weak signature in the sensor measurements. A global analysis approach, that detects broad changes in the generated data, is needed to reveal the incipient fault signature, which is masked by noise. It requires historical data generated by the process in healthy and faulty conditions to be available.

A global time-domain analysis approach applied to ECT signals has proven its efficiency in the detection and characterisation of minor cracks in conductive plates. It is based on KL divergence, which is used as a global fault indicator showing high sensitivity to the small changes in sensor impedance caused by the presence of cracks. The PCA applied to the statistical parameters of the ECT signals allows to obtain a characterisation subspace with a trivial modelling effort.

A global frequency-domain approach has been shown effective to differentiate among different types and sizes of electrical machine faults, specifically bearing faults. The approach overcomes the prior knowledge of characteristic faults frequencies. High-level features have been extracted using the PCA applied to the overall spectral signature of faults. The LDA allows for improving the discrimination results of faults in bearing balls. The two techniques lead to classification subspaces that perform faults discrimination according to the faults types and sizes irrespectively to the machine operating point.

Bibliography

[1] M. A. T. Oral Buyukozturk, *Nondestructive Testing of Materials and Structures*, ser. RILEM Bookseries. Springer Netherlands, 2013, vol. 6.

[2] T. Stepinski, T. Uhl, and W. Staszewski, *Advanced Structural Damage Detection: From Theory to Engineering Applications*. John Wiley & Sons, Ltd., 2013.

[3] M. Friswell and J. E. Mottershead, *Finite element model updating in structural dynamics*. Springer Science & Business Media, 1995.

[4] S. W. Doebling, C. R. Farrar, and M. B. Prime, "A summary review of vibration-based damage identification methods," *Identification Methods," The Shock and Vibration Digest*, vol. 30, pp. 91–105, 1998.

[5] N. Martin, "Advanced signal processing and condition monitoring," *International Journal Insight on Non-Destructive Testing & Condition Monitoring*, vol. 49, no. 8, 2007.

[6] J. P. Santos, C. Cremona, A. D. Orcesi, and P. Silveira, "Multivariate statistical analysis for early damage detection," *Engineering Structures*, vol. 56, pp. 305–335, 2013.

[7] H. J. Lim, M. K. Kim, H. Sohn, and C. Y. Park, "Impedance based damage detection under varying temperature and loading conditions," *NDT & E International*, vol. 44, no. 8, pp. 740–750, 2011.

[8] I. Trendafilova, "An automated procedure for detection and identification of ball bearing damage using multivariate statistics and pattern recognition," *Mechanical Systems and Signal Processing*, vol. 24, no. 6, pp. 1858–1869, 2010.

[9] N. Dervilis, M. Choi, S. G. Taylor, R. J. Barthorpe, G. Park, C. R. Farrar, and K. Worden, "On damage diagnosis for a wind turbine blade using pattern recognition," *Journal of Sound and Vibration*, vol. 333, no. 6, pp. 1833–1850, 2014.

[10] D. Horn and W. R. Mayo, "Nde reliability gains from combining eddy-current and ultrasonic testing," *NDT & E International*, vol. 33, no. 6, pp. 351–362, 2000.

[11] I. Komura, T. Hirasawa, S. Nagai, J. ichi Takabayashi, and K. Naruse, "Crack detection and sizing technique by ultrasonic and electromagnetic methods," *Nuclear Engineering and Design*, vol. 206, no. 2-3, pp. 351–362, 2001.

[12] F. Bos and S. B. Casagrande, "On-line non-destructive evaluation and control of wood-based panels by vibration analysis," *Journal of Sound and Vibration*, vol. 268, no. 2-3, pp. 403–412, 2003.

[13] W. Fan and P. Qiao, "Vibration-based damage identification methods: A review and comparative study," *Structural Health Monitoring*, vol. 10, pp. 83–111, 2011.

[14] M. Njiki, A. Elouardi, S. Bouaziz, O. Casula, and O. Roy, "A real-time implementation of the total focusing method for rapid and precise diagnostic in non destructive evaluation," *IEEE 24th International Conference on Application-Specific Systems, Architectures and Processors (ASAP)*, pp. 245–248, 2013.

[15] M. Sutcliffe, M. Weston, B. Dutton, P. Charlton, and K. Donne, "Real-time full matrix capture for ultrasonic non-destructive testing with acceleration of post-processing through graphic hardware," *IEEE 24th International Conference on Application-Specific Systems, Architectures and Processors (ASAP)*, vol. 51, pp. 16–23, 2012.

[16] J. Garcia-Martin, J. Gomez-Gil, and E. Vazquez-Sanchez, "Non-destructive techniques based on eddy current testing," *Sensors*, vol. 11, no. 3, pp. 2525–2565, 2011.

[17] H. Huang, T. Takagi, and H. Fukutomi, "Fast signal predictions of noised signals in eddy current testing," *IEEE Transactions on Magnetics*, vol. 4, no. 36, pp. 1719–1723, 2000.

[18] C. Huang, W. Xinjun, X. Zhiyuan, and Y. Kang, "Pulsed eddy current signal processing method for signal denoising in ferromagnetic plate testing," *NDT & E International*, vol. 43, no. 7, pp. 648–653, 2010.

[19] E. Cardelli, A. Faba, and F. Tissi, "Contact-less speed probe based on eddy currents," *IEEE Transactions on Magnetics*, vol. 49, no. 7, pp. 3897–3900, 2013.

[20] L. A. N. M. Lopez, D. K. S. Ting, and B. R. Upadhyaya, "Removing eddy-current probe wobble noise from steam generator tubes testing using wavelet transform," *Progress in Nuclear Energy*, vol. 50, no. 7, pp. 828–835, 2008.

[21] P. Henriquez, J. B. Alonso, M. A. Ferrer, and C. M. Travieso, "Review of automatic fault diagnosis systems using audio and vibration signals," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 5, pp. 642–652, 2014.

[22] K. Elbhbah and J. K. Sinha, "Vibration-based condition monitoring of rotating machines using a machine composite spectrum," *Journal of Sound and Vibration*, vol. 332, no. 11, pp. 2831–2845, 2013.

[23] C. Ravat, P. Joubert, Y. L. Bihan, C. Marchand, M.Woytazik, and E. Dufour-Gergam, "Non-destructive evaluation of small defects using an eddy current microcoil sensor array," *Sensor Letters*, vol. 7, no. 3, pp. 400–405, 2009.

[24] N. V. Nair, V. R. Melapudi, H. R. Jimenez, X. Liu, Y. Deng, Z. Zeng, L. Udpa, T. J. Moran, and S. S. Udpa, "A GMR-based eddy current system for NDE of aircraft structures," *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 3312–3314, 2006.

[25] P.-Y. Joubert, Y. L. Bihan, and D. Placko, "Localization of defects in steam generator tubes using a multi-coil eddy current probe dedicated to high speed inspection," *NDT & E International*, vol. 35, no. 1, pp. 53–59, 2002.

[26] G. V. Drunen and V. S. Cecco, "Recognizing limitations in eddy-current testing," *NDT International*, vol. 17, no. 1, pp. 1–6, 1984.

[27] H. Wang, B. Ju, W. Li, and Z. Feng, "Ultrastable eddy current displacement sensor working in harsh temperature environments with comprehensive self-temperature compensation," *Sensors and Actuators A: Physical*, vol. 211, pp. 98–104, 2014.

[28] Y. Yu, Y. Yan, F. Wang, G. Tian, and D. Zhang, "An approach to reduce lift-off noise in pulsed eddy current nondestructive technology," *NDT & E International*, vol. 63, pp. 1–6, 2014.

[29] Y. L. Bihan, J. Pavo, and C. Marchand, "Study and experimental validation of the calculation of the ECT signal induced by a minute crack using a FEM–BIM combination," *NDT & E International*, vol. 39, no. 6, pp. 476–486, 2006.

[30] ——, "Characterization of small cracks in eddy current testing," *The European Physical Journal Applied Physics*, vol. 43, no. 2, pp. 231–237, 2008.

[31] T. Boukra, A. Lebaroud, and G. Clerc, "Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 9, pp. 4034–4042, 2013.

[32] S. Nandi and H. Toliyat, "Condition monitoring and fault diagnosis of electrical motors-a review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, pp. 719–729, 2005.

[33] M. E. H. Benbouzid, "Bibliography on induction motors faults detection and diagnosis," *IEEE Transactions on Energy Conversion*, vol. 14, no. 54, pp. 1065–1074, 1999.

[34] T. Harris, *Rolling Bearing Analysis.* 3rd Ed. New York: Wiley, 1991.

[35] EPRI, "Improved motors for utility applications," *Publication EL-2678-V1, final report*, 1982.

[36] O. V. Thorsen and M. Dalva, "Failure identification and analysis for high voltage induction motors in the petrochemical industry," *IEEE Transactions on Industry Applications*, vol. 35, no. 4, pp. 810–818, 1999.

[37] S. Barker, "Avoiding premature bearing failure with inverter fed induction motors," *Power Engineering Journal*, vol. 14, no. 4, pp. 182–189, 2000.

[38] I. Y. Onel and M. El Hachemi Benbouzid, "Induction motor bearing failure detection and diagnosis: Park and Concordia transform approaches comparative study," *IEEE/ASME Transactions on Mechatronics*, vol. 13, no. 2, pp. 257–262, 2008.

[39] E. H. E. Bouchikhi, V. Choqueuse, and M. E. H. Benbouzid, "Current frequency spectral subtraction and its contribution to induction machines' bearings condition monitoring," *IEEE Transactions on Energy Conversion*, vol. 28, no. 1, pp. 135–144, 2013.

[40] B. Trajin, J. Regnier, and J. Faucher, "Comparison between vibration and stator current analysis for the detection of bearing faults in asynchronous drives," *Electric Power Applications, IET*, vol. 4, no. 2, pp. 90–100, 2010.

[41] M. Tsypkin, "Induction motor condition monitoring: Vibration analysis technique - a twice line frequency component as a diagnostic tool," *IEEE International Electric Machines & Drives Conference (IEMDC)*, vol. 4, no. 2, pp. 117–124, 2013.

[42] L. Frosini and E. Bassi, "Stator current and motor efficiency as indicators for different types of bearing faults in induction motors," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 1, pp. 244–251, 2010.

[43] S. Ahmed and G. Singhn, "Experimental investigations on induction machine condition monitoring and fault diagnosis using digital signal processing techniques," *Electric Power System Research*, vol. 65, no. 3, pp. 197–221, 2003.

[44] G. Dong and J. Chen, "Noise resistant time frequency analysis and application in fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 33, pp. 212–236, 2012.

[45] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using fft of intrinsic mode functions in hilbert–huang transform," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2607–2615, 2007.

[46] R. Randall and J. Antoni, "Rolling element bearing diagnostics - a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.

[47] R. Collacott, *Vibration Monitoring and Diagnosis*. New York: John Wiley & Sons, 1979.

[48] K. Worden, W. Staszewski, and J. Hensman, "Natural computing for mechanical systems research: A tutorial overview," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 4–111, 2011.

[49] R. Rubini and U. Meneghetti, "Application of the envelope and wavelet transform analysis for the diagnosis of incipient faults in ball bearings," *Mechanical Systems and Signal Processing*, vol. 15, no. 2, pp. 28–302, 2001.

[50] H. Li, H. Zheng, and L. Tang, "Wigner-Ville Distribution based on EMD for faults diagnosis of bearing," *Lecture Notes in Computer Science*, vol. 4223, pp. 803–812, 2006.

[51] R. Klein, D. Ingman, and S. Braun, "Non-stationary signals: phase-energy approach-theory and simulations," *Mechanical Systems and Signal Processing*, vol. 15, no. 6, pp. 1061–1089, 2001.

[52] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. E. Orchard, and G. Vachtsevanos, "A probabilistic fault detection approach: Application to bearing fault detection," *IEEE Transactions on Industry Applications*, vol. 58, no. 5, pp. 2011–2018, 2011.

[53] Y. Wang, Z. He, and Y. Zi, "Enhancement of signal denoising and multiple fault signatures detecting in rotating machinery using dual-tree complex wavelet transform," *Mechanical Systems and Signal Processing*, vol. 24, no. 1, pp. 119–137, 2010.

[54] P. McFadden and J. Smith, "Vibration monitoring of rolling element bearings by the high frequency resonance technique - a review," *Tribology International*, vol. 17, no. 1, pp. 3–10, 1984.

[55] R. B. Randall, J. Antoni, and S. Chobsaard, "A comparison of cyclostationary and envelope analysis in the diagnostics of rolling element bearings," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, pp. 3882–3885, 2000.

[56] D. Siegel, C. Ly, and J. Lee, "Methodology and framework for predicting helicopter rolling element bearing failure," *IEEE Transactions on Reliability*, vol. 61, no. 4, pp. 846–857, 2012.

[57] G. Feng, A. Mustafa, J. X. Gu, D. Zhen, F. Gu, and A. D. Ball, "The real-time implementation of envelope analysis for bearing fault diagnosis based on wireless sensor network," *19th International Conference on Automation and Computing (ICAC)*, pp. 1–6, 2013.

[58] S. Hahn, *Hilbert transforms in signal processing.* Artech House, 1996.

[59] F. Cong, J. Chen, G. Dong, and M. Pecht, "Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis," *Journal of Sound and Vibration*, vol. 332, no. 8, pp. 2081–2097, 2013.

[60] J. I. Taylor, *The Vibration Analysis Handbook.* Vibration Consultants, 1994.

[61] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and genetic algorithm for bearing fault detection," *Soft Computing*, vol. 10, no. 3, pp. 264–271, 2006.

[62] B. Lazzerini and S. L. Volpi, "Classifier ensembles to improve the robustness to noise of bearing fault diagnosis," *Pattern Analysis and Applications*, vol. 16, no. 2, pp. 235–251, 2013.

[63] F. Immovilli, M. Cocconcelli, A. Bellini, and R. Rubini, "Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 11, pp. 4710–4717, 2009.

[64] T. Jing, C. Morillo, and M. G. Pecht, "Rolling element bearing fault diagnosis using simulated annealing optimized spectral kurtosis," *IEEE Conference on Prognostics and Health Management (PHM)*, pp. 1–5, 2013.

[65] D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 176–199, 2013.

[66] J. Antoni and R. B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machine," *Mechanical System and Signal Processing*, vol. 20, no. 2, pp. 308–331, 2006.

[67] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. New York: Academic, 1979.

[68] J. Harmouche, C. Delpha, and D. Diallo, "Linear discriminant analysis for the discrimination of faults in bearing balls by using spectral features," *International Conference on Green Energy*, 2014.

[69] "Case western reserve university, bearing data centre," "URL:http://csegroups.case.edu/bearingdatacenter/pages/download-data-file"/, 2014, [Online; accessed 13-June-2014].

[70] B. Samanta and K. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features," *Mechanical Systems and Signal Processing*, vol. 17, no. 2, pp. 317–328, 2003.

[71] J. Zarei, "Induction motors bearing fault detection using pattern recognition techniques," *Expert Systems with Applications*, vol. 39, no. 1, pp. 68–73, 2012.

[72] J. Harmouche, C. Delpha, and D. Diallo, "Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals," *IEEE Transactions on Energy Conversion*, to be published.

[73] C. Choi and C. Lee, "Feature extraction based on the bhattacharyya distance," *Pattern Recognition*, vol. 36, pp. 1703–1709, 2003.

[74] C. Lee and E. Choi, "Bayes error evaluation of the gaussian ml classifer," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 1471–1475, 2000.

[75] B. Muruganatham, M. A. Sanjith, B. Krishnakumar, and S. S. Murty, "Roller element bearing fault diagnosis using singular spectrum analysis," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 150–166, 2013.

Conclusions and perspectives

The fault detection and diagnosis is a basis function of condition-based monitoring system. An efficient FDD algorithm that provides early warnings of faults and identifies the fault type/location and severity is essential to make the appropriate maintenance and control measures. The result is an increase in the availability, reliability and safety of industrial processes.

The sensor(s) data which can be of different types, i.e. electrical, mechanical, acoustic, thermography, etc. are generally processed through transformations that reveal the faults signature in form of artificial signals, i.e. 'residuals', or features. These transformations can be either quantitative and/or qualitative model-based, or data-driven using advanced signal processing techniques and multivariate analysis. Data-driven approaches which do not require any form of signal/process model are especially useful when it comes to large-scale and complex multi-physics systems with coupled and non-linear phenomena. This is the case of most processes involved in industry and transportation, where the *a priori* knowledge available about the process dynamics is usually not sufficient for building an accurate explicit model. Data-driven approaches require a minimum of prior knowledge and modelling effort. They allow for an efficient fault detection and diagnosis, providing that a good historical database, which covers all operating points and fault scenarios, is available. In the context of data-driven approaches, the problem of FDD was addressed for the special case of incipient faults, and for the particular case where they are buried in noise.

Real faults at incipient stage cause imperceptible changes that range near the noise level ($FNR \leq 0$ dB) in the process measurements. The statistical process monitoring uses statistical control charts that are basically designed to detect shifts in parameters of the process probability distribution. The detection of incipient faults, however, requires the fault indicator to be sensitive to small general changes: in real applications, the process data can have nondescript probability distributions and the incipient fault signature is weak and unpredictable. In this distribution-free and non-parametric framework, the Kullback-Leibler divergence was proposed in

the second part of the study to detect the global change caused by incipient faults. Using numeral examples, the divergence proved to have a better detection performance than the one obtained with the $\mathbf{T^2}$ and the SPE. The latter statistics are usually used for fault detection with the multivariate subspace projection methods, but compared to the divergence they showed poor capability to detect the short duration change (10% of sample size) whose amplitude is less than 10% of the signal magnitude. For such faults and favourable conditions of noise ($SNR = 35$ dB), the divergence achieved zero probabilities of detection errors. In noisy environments ($SNR \leq 25$ dB), the divergence sensitivity with respect to incipient faults can be readily enhanced by computing the divergence on large data samples.

In the third part of this work, the fault amplitude estimation using the KL divergence was addressed. In PCA-based data representation and under some assumptions of data and fault modelling, a theoretical study was conducted to build an analytical model of the divergence that is a function of the PCA's parameters and the fault amplitude. A theoretical fault amplitude estimator that depends on the divergence value was obtained. The estimator, when evaluated on numerical examples, gave an overestimation of the actual fault amplitude, guaranteeing a safety margin for the monitored process. The estimation accuracy was evaluated for incipient faults in both favourable and worst conditions of noise level and data size. In worst case, the relative error obtained on the estimation of the faulty variable did not exceed 1%.

As an alternative to the 'global' monitoring of probability distributions in time-domain signals, a 'global' analysis of the signals spectra was shown to be relevant in diagnosing incipient faults having a spectral signature. The 'global' approach was validated by application to non-destructive inspection-based damage detection. Many engineering systems are subject to faults that start with small material cracks, but can rapidly develop into a serious damage leading to a catastrophic failure. Ensuring the integrity of structures is crucial for safety-critical applications like aeroplanes, ships, submarine, nuclear reactors, etc.

A first application study was dedicated to the detection and characterisation of small surface cracks in conductive structure using experimental ECT data. The detection of the smallest cracks, typically the cracks with length=$\{0.1$ mm,0.2 mm$\}$ and depth=0.1 mm, was challenging: the change in the probe impedance due to the crack presence was very weak and concealed by the noise. The disparity between the pdfs of the reference ECT signal and the faulty one was small enough so that the first statistical moments failed to detect the difference. The KL divergence, in contrast, was able to reveal the dissimilarity, making small cracks detection successful. In

addition, a characterisation subspace for crack severity assessment was built based on the statistical analysis of the ECT signals. PCA when applied to several statistical features, including the divergence, gave two principal components which are sensitive to the length and surface of cracks.

In a second study, the detection and diagnosis of bearing faults in electrical rotating machines using vibration signals was of concern. A 2-dimensional PCA subspace for fault classification was obtained based on the analysis of the 'global' spectral signature of faults. Except for the discrimination among ball fault sizes, the PCA subspace was shown able to discriminate among different types (BF, ORF and IRF) and sizes (2 small sizes for each fault) of bearing faults. The detection of ball faults is known to be the most critical and difficult compared to the other bearing fault types, since the associated signature is usually very weak and masked by the high noise level of vibration signals. However, the LDA when it was applied to the dominant spectral features of the vibration spectra, contributed to improving the discrimination results of faults in bearing balls. The two multivariate techniques led to classification subspaces that allow the discrimination among the faults types and sizes without being related to the machine operating point. In addition, the 'global' frequency-domain analysis is less dependent on the characteristic fault frequencies, their exact positions and prior knowledge.

Concerning future works, both theoretical issues and applications could be addressed.

Despite the promising results for fault detection and identification in ECT, there is still an issue for very small cracks with surfaces less than 100 mm^2. The introduction of analytical knowledge in the classification step would improve the characterisation capability. Moreover, it would be useful to develop an analytical model for the fault identification or characterisation from the experimental data. The method will then be easier to implement for real-time embedded applications.

Furthermore, data driven methods have proven to be efficient for large distributed systems. KL divergence should be evaluated for fault diagnosis in microgrids, in offshore networks where access to all measurements is tedious or impossible and in transportation applications. To enhance the theoretical contributions, one can:

- improve the performance (e.g. missed detection and false alarm probabilities) of the fault detection by the optimisation of the detection threshold
- enhance the evaluation of the influence of the different parameters (data size, detection threshold, SNR, ...) on the KL divergence fault detection capability

- extend the analytical model for fault severity estimation to other types of probability distribution
- extend the KL divergence capability to prognosis with the use of dynamic multi-dimensional data representation
- develop a hybrid approach based on the judicious combination of data-driven and analytical fault diagnosis problem formulations

Appendix

A1. MATLAB functions

% X: reference data matrix $[m \times N]$
% Xf: faulty data matrix $[m \times N]$
% S: covariance matrix $[m \times m]$
[m,N]=size(X)
aveX=mean(X);
Xst=X-aveX;
S=Xst'*Xst/N;
P, eigen = eigs(S,m);
T=Xst*P;

% normalised reference pdf of the kth principal component T(:,k):
[h,u]=ksdensity(T(:,k));
h = h./sum(h);

% normalised faulty pdf of the kth principal component Tf(:,k):
Xfst=(Xf-aveX);
Tf=Xfst*P;
q=ksdensity(Tf(:,k),u);
q = q./sum(q);

% KL divergence approximation:
$\hat{D}$=sum((h-q).*log(h./q));

% The divergence threshold for a particular probability of false alarm
The divergence threshold is calculated irrespectively of the fault characteristics, but from training (fault-free) data, according to the procedure described in Fig.4.26.

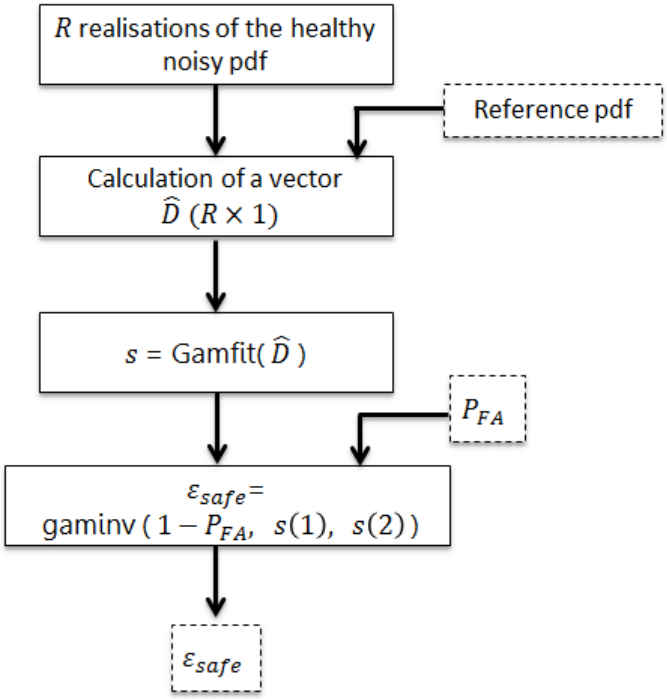

Figure 4.26: Divergence threshold ϵ_{safe} calculation for a particular P_{FA}

A2. Setting parameters of the divergence measure

Three parameters are needed to approximate the divergence measure using the empirical probabilities:

- the number n_I of the disjoint intervals or bins or equally spaced points
- the type of the applied kernel function
- the bandwidth of the kernel window

The statistics toolbox of MATLAB provides the user with the function ksdensity(**x**) that estimates the probability density for a vector **x** at n_I equally spaced points using a kernel smoothing function, that can be normal, box, epanechnikov or triangle. The function hist(**x**) calculates the histogram of **x**, from which the probability distribution histogram can be estimated. The latter corresponds to the application of a rectangle kernel in the kernel density estimation. The calculation of the divergence through probability distribution histograms is faster than via kernel density estimates, but the kernel density estimate is more robust with respect to noise, which is important for the detection purpose. To illustrate the effect of the kernel type on the detection performance, we evaluate the missed detection probability P_{MD}

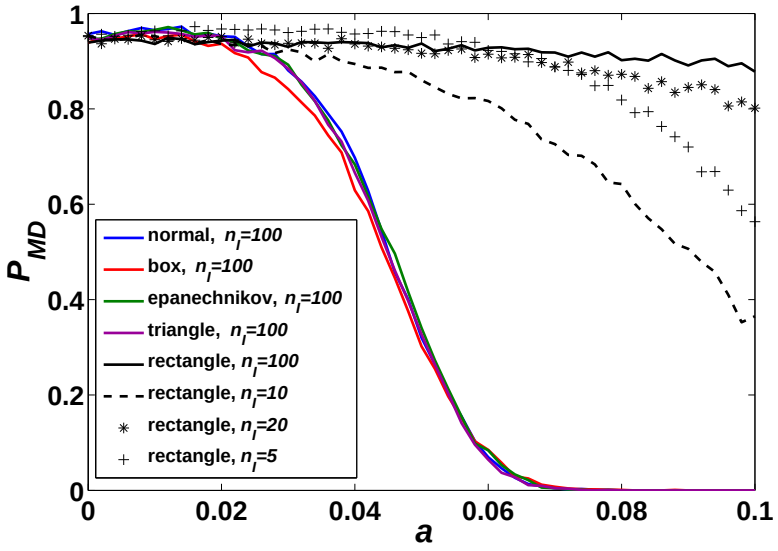

Figure 4.27: P_{MD} versus the fault amplitude a, different kernels

on the numerical process used in chapter 1 (reminder: the variable x_1 is faulty, the divergence is calculated on the first principal component t_1). The evaluation is carried out in the following conditions: the data length $N = 1000$, the $SNR = 15dB$ and the N observations are supposed faulty. Fig.4.27 is obtained. The divergence calculated via histograms has poor detection performance compared to its counterpart that uses kernel density estimates. The other kernels have similar detection performance. This result can be interpreted by the fact that the divergence performs a relative comparison between probability distributions, and thus an accurate representation of the underlying distributions is not necessary for the purpose of fault detection. The P_{MD} is also evaluated for different number of disjoint bins (n_I), in case of normal kernel density estimation. Fig.4.28 is obtained, and it shows that a too low number of bins (n_I=10 when N=1000) can degrade the detection performance. Otherwise, any reasonable number of bins can be used. The results that were presented are obtained with a normal kernel and $n_I = 100$.

The kernel bandwidth is the smoothness factor of the probability density estimate. The optimal bandwidth depends on the true underlying density, and the common method to choose the optimal bandwidth is to use the bandwidth that minimises the AMISE (Asymptotic Mean Integrated Squared Error) criterion. The default bandwidth h_0 used by ksdensity is optimal for estimating normal densities. The function allows however choosing another value. It is worth reminding that the problem of fault detection using KL divergence is not related to the problem of accurately estimating the underlying probability densities. The setting parameter

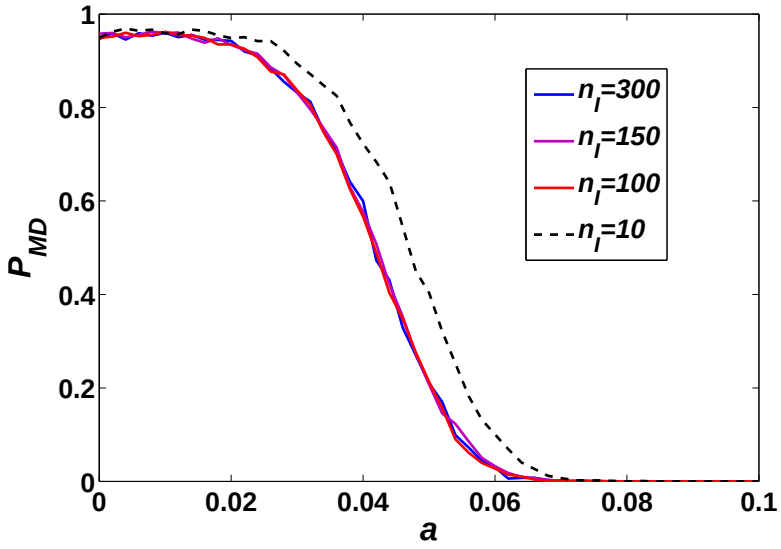

Figure 4.28: P_{MD} versus the fault amplitude a, normal kernel

values that maximise the sensitivity of the divergence to small faults and its robustness to noise are not necessarily those which accurately estimate the pdf. Fig.4.29

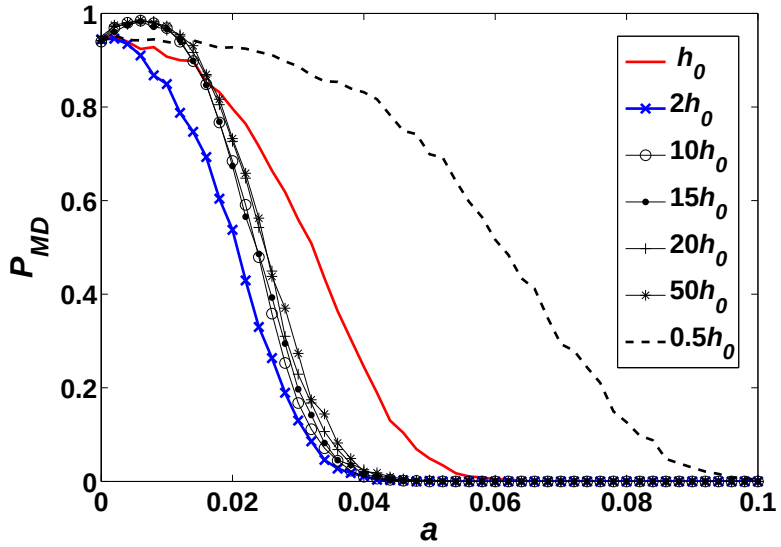

Figure 4.29: P_{MD} versus the fault amplitude a, different normal kernel bandwidths

shows the probability of missed detection, for different values of the normal kernel bandwidth. In general, small kernel bandwidth values lead to undersmoothed estimates, while large values lead to oversmoothing. With increasing bandwidth, the pdfs get less sensitive to the random variation in the random data, thus increasing

the sensitivity to the small deterministic variation due to faults. That is how we interpret the behaviour shown in Fig.4.29. So high values of the kernel bandwidth are recommended for the purpose of fault detection. However, all the results presented in this dissertation were obtained with the default bandwidth h_0 : the detection and estimation performance were satisfactory even with a non-optimal parameters setting.