IntechOpen

Fault Diagnosis and Detection

Edited by Mustafa Demetgul and Muhammet Ünal

FAULT DIAGNOSIS AND DETECTION

Edited by **Mustafa Demetgul**
and **Muhammet Ünal**

Fault Diagnosis and Detection
http://dx.doi.org/10.5772/63169
Edited by Mustafa Demetgul and Muhammet Ünal

Contributors

Mircea Ruba, Miguel Delgado-Prieto, Daniel Zurita-Millán, Jesus A. Cariño, Marta Millán Gonzalvez, Juan A. Ortega Redondo, René De J. Romero Troncoso, Yu Liu, Deepshikha Agarwal, Muhammad Irfan, Nordin Bin Saad, Rosdiazli B. Ibrahim Ibrahim, Vijanth Sagayan, Abdullah Alwadie, Muhammad Aman Sheikh, Mohamed N. Nounou, M. Ziyan Sheriff, Chiranjivi Botre, Majdi Mansouri, M. Nazmul Karim, Hazem Numan Nounou, Dusan Krokavec, Anna Filasova, Pavol Liscinsky, Xihui Liang, Efrain Alcorta-Garcia, Luis Humberto Rodriguez Alfaro, Cornelio Posadas Castillo, David Alejandro Diaz Romero, Minescu Mihail, Pana Ion, Lahouari Cheded, Rajamani Doraiswami, Wei Guo, Ming J. Zuo, Wenhua Chen, Nursyarizal Mohd Nor, Taib Ibrahim

Notice

First published in Croatia, 2017 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Fault Diagnosis and Detection
Edited by Mustafa Demetgul and Muhammet Ünal

p. cm.

Print ISBN 978-953-51-3203-5

Online ISBN 978-953-51-3204-2

eBook (PDF) ISBN 978-953-51-4822-7

Meet the editors

Mustafa Demetgul was born in Duzce, Turkey, in 1978. He received the BSc degree, MS degree, and PhD degree in Mechanical Engineering from Marmara University, Istanbul, Turkey. From 2007 to 2009, he was a postdoc with the Mechanical Engineering Department, Florida International University, Florida, USA. Here he has been involved in research on structural health monitoring and energy harvesting. He currently works at the Department of Mechatronics Engineering, Marmara University, as an Assoc. Prof. Dr. His research interests are fault diagnosis, energy harvesting, intelligent materials, automation systems, structural health monitoring, and artificial intelligence systems. He is the director of Fault Diagnosis and Structural Health Monitoring Lab in Marmara University. He is an editorial board member at many international journals. He has publications in many international journals and symposiums on failure diagnosis and fault detection.

Dr. Muhammet Ünal works as an assistant professor in the Department of Mechatronics Engineering, Marmara University, Istanbul (Turkey). His research interests are system identification, parameter optimization, real-time control, and structural health monitoring. His primary research interests are vibration analysis and fault diagnosis of mechanical systems. Dr. Ünal has written several papers in vrious fields of mechatronics engineering and one book on control engineering. He is also the first author of the book Optimization of PID Controllers Using Ant Colony and Genetic Algorithms, which is published in Springer.

Contents

Preface

Mass production companies have become obliged to reduce their production costs and sell more products with lower profit margins in order to survive in competitive market conditions. The complexity and automation level of machinery are continuously growing. This development calls for some of the most critical issues that are reliability and dependability of automatic systems. Furthermore, the maintenance and repair of these machines are both costly and troublesome. In the factories where such automation systems are in use, even a small fault in the production band stops the entire production line. Fault is defined as the status where at least one characteristic feature or parameter is outside the limits allowed in the normal state. These faults degrade the performance of the system by either causing damage to the system or completely disrupting the system. In production plants, even a few minutes of production stop results in huge production losses and a remarkable increase in costs. These unexpected failures take a long time to be repaired. The factories are trying to pass through conventional maintenance and fault detection methods using various tests and analysis after a fault occurs. In particular, car factories reserve a big portion of their budgets to provide resources and trained maintenance personnel for detecting and solving any such faults as quickly as possible. Since industrial maintenance costs account for 4–25% of the total cost of production, technical maintenance departments have become one of the most important parts in large factories.

In many production plants, machines are diagnosed and maintained with predictive maintenance methods or trained maintenance personnel. Predictive maintenance is in the last place in maintenance variety with 12% in practice. However, predictive maintenance provides many opportunities, which bring many opportunities for businesses. Nevertheless, predictive maintenance work can be carried out in certain periods. Failures that cannot be foreseen during such a maintenance are detected after occurrence of the failure. These abnormal events have significant economic, safety, and environmental impacts. Fault diagnosis is mostly possible with predictive maintenance instruments. The main disadvantages of these machines are the cost of the instruments and requirement of experienced personnel for interpreting the data. Moreover, even large-scale companies are only able to check their machines once a failure happens or at prescheduled intervals. However, this is not a healthy machine monitoring approach. In contrast, the ideal approach is monitoring the machines online and handling fault diagnosis with artificial intelligence techniques. In this context, machines will be monitored remotely, and computer-aided techniques will be employed to detect faults in the future. Monitoring of status of machines with such artificial intelligence will be provided by certain companies as an application service in cloud computing.

Depending on the advancement of technology, many factories are controlled by unmanned intelligent systems. Recently, the concept Industry 4.0 gained great importance in Europe.

Current practices and studies focus almost exclusively on better products. Some of the most apparent aims of Industry 4.0 are factory monitoring and predictive maintenance (production performance analysis, machine status, abnormal diagnostic, and sensor measurement). Research on this issue is still ongoing. In South Korea, some manufacturers of car parts are remotely controlling their production activities. But there is not a factory where all the machines are solely controlled remotely.

A plant or system consists of three subsections: actuators, main process, and sensors. Over time, there occur faults due to environmental reasons, aging, production mistakes, inappropriate working conditions, misuse, human errors, etc. Fault diagnosis techniques are composed of three steps: fault detection, fault isolation, and fault analysis and identifications. Fault detection is used to check whether there is any fault in the system and determine as soon as it occurs. As the fault is detected, fault isolation is used to identify the location of the faulty component. Then, fault analysis and identification are used to specify fault features (e.g., type, shape, and size). A number of techniques are used in fault diagnosis. These techniques can be categorized as mathematical-based models and knowledge-based and data-driven ones. Many factors such as process types, unexpected disturbances, open- or close-loop structure, process nonlinearities, etc. influence the effectiveness and correctness of the diagnosis.

Fault Diagnostics are used in many different industrial sectors like space systems, defense systems (aviation, marine, and ground), commercial aerospace (aircraft, jet engines), ground vehicles (locomotives, trucks, cars), high tech (networks and IT systems , disk drives, server farms), and process control (IC manufacturing, refineries , power plants, oil and gas drilling).

We invited selected authors to contribute their original research articles as well as provide literature review articles that will illustrate and stimulate the continuing effort to understand the fault diagnosis and detection. In this book, there are sections on bearing failures, gear failures, electric motors, wind farms, petroleum equipment, planetary gearboxes, induction motors, rotating machinery, and electromechanical system fault diagnosis. The accepted papers show a diversity of new developments in these areas. This issue has high-quality articles containing original research results and survey articles of exceptional merit, and it will let the readers of this book know more about fault diagnosis and detection.

In the future, there will be unmanned factories where machines and systems communicate to each other, detect their own faults, and can remotely intercept their faults. The pioneer studies of such systems are fault diagnosis studies. Thus, we hope that this book will contribute to the literature in this regard.

Dr. Mustafa Demetgul and **Dr. Muhammet Ünal**
Department of Mechatronics Engineering
Marmara University
Istanbul, Turkey

Chapter 1

Fault Detection and Isolation

Rajamani Doraiswami and Lahouari Cheded

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67870

Abstract

Fault diagnosis of a class of linear multiple-input and multiple-output (MIMO) systems is developed here. An emulator-based scheme is proposed to detect and isolate faults in a system formed by interconnected subsystems. Emulators, which are hardware or software devices, are connected to the input and measurement outputs in cascade with the subsystems whose faults are to be diagnosed. The role of an emulator is to induce variations in cascade combination of the nominal fault-free subsystem so as to mimic the actual perturbations that may occur in the subsystem during the offline identification phase. The emulator-generated data are employed in the reliable identification of the nominal system, the associated Kalman filter, and a map that relates the emulator parameters to the feature vector. In the operational stage, the Kalman filter residual is used to detect a fault in the system; the emulator parameter that has varied is estimated, and using the emulator-feature vector map, the faulty subsystem is isolated. The main contributions of this work are accurate and reliable identification of the system, the fault diagnosis of multivariable systems using feature vector-emulator map fault diagnosis of multivariable systems, and the establishment of the key properties of the Kalman filter for fault detection. The proposed scheme was successfully evaluated on a number of simulated as well as physical systems.

Keywords: fault detection, fault isolation, fault diagnosis, Kalman filter, emulators, identification, Bayes decision theory

1. Introduction

Fault detection and isolation (FDI) of physical systems—especially mission critical systems including nuclear reactors, aircraft, automotive systems, spacecraft, autonomous vehicles, and fast rail transportation—is becoming increasingly important in recent times thanks mainly to advances in sensors, computing, and communication technologies. It still poses a challenge in

view of the stringent and conflicting requirements, high probability of correct detection and isolation, low false alarm probability, and timely decision on the fault status.

The identification of the system model is crucial to the performance of the fault diagnosis scheme. The more accurate the identified model, the higher is the probability of correct diagnosis and the lower is the false alarm probability. The reliability and accuracy of the identification hinges on ensuring that the identified model is captured completely and what is leftover is the information-less zero-mean white noise process. As the Kalman filter is a zero-mean white noise process if and only if there is no mismatch between the identified model and the model of the system, the identification scheme should minimize the residual of the Kalman filter—instead the equation error, which in general, is a colored noise [1]. The widely popular, consistent, and efficient scheme that meets the above state requirement is the prediction error method (PEM) [2]. The PEM identifies the system by minimizing the residual of the Kalman filter.

A physical system is subject to perturbation resulting from the variations of the parameters and effects nonlinearities resulting in the deviation in the neighborhood of the nominal operating point. A model identified at a nominal operating point will not capture the static and the dynamic behavior of the perturbed system. To overcome this, an emulator, which is a hardware or a software device, is connected to either an accessible input or an accessible output in cascade with a subsystem to mimic its operating scenarios [3–5]. The powerful concept of emulators, which is employed to mimic the likely operating scenarios for single-input and single-output (SISO) system, is extended to multiple-input and multiple-output (MIMO) and multiple-input and single-output (MISO) system. The system is identified and the feature vector-emulator map is estimated from the emulator-generated data covering all likely operating scenarios including the normal and the faulty ones similar in spirit to that employed in training the neural network [6]. The identified nominal model, an *optimal nominal model*, is robust to model perturbation in the neighborhoods of the nominal operating point. It may be worth noting that the conventional scheme uses only the input-output data from the system in the nominal operating scheme.

There are essentially three approaches to the failure detection and isolation problem: the non-parametric approach, the parametric approach, and the combined approach. The non-parametric approach is based on analyzing a residual. The residual is defined as a signal, which is ideally non-zero in a statistical sense when there is a failure present, and zero otherwise. The residual may be generated using Kalman filters, observers, unknown-input observers, other forms of detection filters, and parity equations [7–12]. In view of the following key properties of the Kalman filter listed below, the Kalman filter is deemed the most preferable for both fault detection and fault isolation [1]:

a. *Model matching*: The residual is a zero-mean white noise process *if and only if* there is no mismatch between the actual model of the system and its identified model embodied in the Kalman filter, that is, and its variance is minimum.

b. *Optimal estimation*: The estimate is optimal in the sense that it is the best estimate that can be obtained by any estimator in the class of all estimators that are constrained by the same assumptions.

c. *Robustness*: Thanks to the feedback (closed-loop) configuration of the Kalman filter with residual feedback, the Kalman filter provides the highest robustness against the effect of disturbance and model variations.

d. *Model mismatch*: If there is a model mismatch, the residual will not be a zero-mean white noise process and an additive term termed fault-indicative term. The fault-indicative term is affine in the deviation in the linear regression or the transfer function model.

The feature vector-emulator map relating the deviation of the feature vector and variations of the emulator parameter is used for fault isolation if a fault is detected. The influence vector, which is the partial derivative of the feature vector with respect to an emulator parameter, plays a crucial role in pinpointing the faulty subsystem and tracks its parameter variation.

The main contributions here are the development of emulator-based system identification, and estimation of the feature vector-emulator map and its application to performance monitoring and fault diagnosis of multivariable system. The key properties of the Kalman filter, including model matching, whitening of the equation error, and residual expression for the model-mismatch case, are established for MIMO, MISO, and SISO systems.

The chapter is organized as follows. In Section 2, the mathematical model of the multiple-input and multiple-output system in state-space, frequency-domain, and a linear regression form is developed. The multiple-input and single-output and the single-input, single-output models are derived. Modeling of faults is also given. In Section 3, the concept of emulators, the generation of emulator-perturbed data, and its role in the identification of the system, the estimation of the feature vector-emulator map for fault isolation is developed. In Section 4, the identification of the system and the associated Kalman filter using prediction error method is suggested. The feature vector-emulator map is estimated using the expression of the Kaman filter residual in the model-mismatch case. In Section 5, the model of the Kalman filter, residual model, and the key properties of this filter are given. The key properties of the residual are established including whitening of the equation error, and expressions for the residual for the model-mismatch case. In Section 6, Bayesian approach to fault diagnosis is explained. Finally, in Sections 7 and 8, the successful evaluation of the proposed scheme on both a simulated and physical system is given, respectively.

2. Mathematical model of the system

The MIMO state-space model of the system denoted $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is given by

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{r}(k) + \boldsymbol{E}_w\boldsymbol{w}(k) \\ \boldsymbol{y}(k) &= \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{v}(k) \end{aligned} \tag{1}$$

where $\boldsymbol{x}(k) = [\, x_1(k) \quad x_2(k) \quad x_3(k) \quad \ldots \quad x_n(k) \,]^T$, $\boldsymbol{y}(k) = [\, y_1(k) \quad y_2(k) \quad y_3(k) \quad \ldots \quad y_q(k) \,]^T$, $\boldsymbol{r}(k) = [\, r_1(k) \quad r_2(k) \quad r_3(k) \quad \ldots \quad r_p(k) \,]^T$, $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$, are respectively, $nx1$ state vector, $qx1$ output, $px1$ input to the system, $px1$ disturbance and $qx1$ measurement noise; $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{E}_w$ are nxn state transition, nxp input, and qxn output and nxp input disturbance matrices; $\boldsymbol{A}$ and $\boldsymbol{C}$ are block

diagonal matrices; $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} & . & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 & . & \boldsymbol{0} \\ . & . & . & . \\ \boldsymbol{0} & \boldsymbol{0} & . & \boldsymbol{A}_q \end{bmatrix}; \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \\ . \\ \boldsymbol{B}_q \end{bmatrix}; \boldsymbol{E}_w = \begin{bmatrix} \boldsymbol{E}_{w1} \\ \boldsymbol{E}_{w2} \\ . \\ \boldsymbol{E}_{wq} \end{bmatrix}; \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{0} & . & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 & . & \boldsymbol{0} \\ . & . & . & . \\ \boldsymbol{0} & \boldsymbol{0} & . & \boldsymbol{C}_q \end{bmatrix}$, $\boldsymbol{A}_j$, $\boldsymbol{B}_j$, $\boldsymbol{E}_{wj,}$and $\boldsymbol{C}_j$ are, respectively, $n_j x n_j$, $n_j x p$, $n_j x p$, and $1 x n_j$ matrices. The output of the system is corrupted by disturbance $\boldsymbol{w}(k)$ and measurement noise $\boldsymbol{v}(k)$; $\boldsymbol{G}(z) = \boldsymbol{C}(z\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B} = D^{-1}(z)$ $\boldsymbol{N}(z)$; $\boldsymbol{I}$ is an identity matrix; $D(z) = |(z\boldsymbol{I} - \boldsymbol{A})| = 1 + \sum_{\ell=1}^{n} a_\ell z^{-\ell}$; $\boldsymbol{B}_j = [\, B_{j1} \quad B_{j2} \quad . \quad B_{jp} \,]$; $\boldsymbol{E}_w = [\, E_{wj1} \quad E_{wj2} \quad . \quad E_{wjp} \,]$.

We assume that the system is controllable and observable, that is, $(\boldsymbol{A}, \boldsymbol{C})$ is observable, $(\boldsymbol{A}, \boldsymbol{B})$ is controllable, implying that all the states may be estimated from the input and the output data, and the input affects all the states. The disturbance $\boldsymbol{w}(k)$ and the measurement noise $\boldsymbol{v}(k)$ are assumed zero-mean white noise processes. The covariance of $w(k)$ and $v(k)$ are

$$E[\boldsymbol{w}\boldsymbol{w}^T] = \boldsymbol{Q} \quad \text{and} \quad E[\boldsymbol{v}\boldsymbol{v}^T] = \boldsymbol{R} \tag{2}$$

where $\boldsymbol{Q}$ and $\boldsymbol{R}$ are positive definite and positive semi-definite matrices, $\boldsymbol{Q} > 0$ and $\boldsymbol{R} \geq 0$. The covariances $\boldsymbol{Q}$ and $\boldsymbol{R}$ are not known a priori.

The MIMO model in the frequency domain is

$$\boldsymbol{y}(z) = \boldsymbol{G}(z)\boldsymbol{r}(z) + \boldsymbol{\vartheta}(z) \tag{3}$$

where $\boldsymbol{G}(z)$ is qxp matrix transfer function, and $\boldsymbol{N}(z)$ is the qxp numerator matrix; $\boldsymbol{\vartheta}(z)$ is the $qx1$ is the effect of disturbance $\boldsymbol{w}(k)$ and the measurement noise $\boldsymbol{v}(k)$ on the output $\boldsymbol{y}(z)$.

$$\boldsymbol{\vartheta}(z) = \boldsymbol{C}(z\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{E}_w\boldsymbol{w}(z) + \boldsymbol{v}(z) \tag{4}$$

2.1. Single-input single-output pairing

A single-input single-output (SISO) model derived from the state-space model relating the input $r_i(z)$, and its associated output, termed $y_{ji}(z)$, which is the same as the output $y_j(z)$ when the input is $r_i(z)$ and the rest of the inputs $r_j(z) = 0$ for $j \neq i$, is

$$y_{ji}(z) = G_{ji}(z)r_i(z) + \vartheta_{ji}(z) \tag{5}$$

where $G_{ji}(z) = \boldsymbol{C}_j(z\boldsymbol{I} - \boldsymbol{A}_j)^{-1}\boldsymbol{B}_{ji} = D_j^{-1}(z)\boldsymbol{N}_{ji}(z)$; and $\vartheta_{ji}(z) = \boldsymbol{C}_j(z\boldsymbol{I} - \boldsymbol{A}_j)^{-1}\boldsymbol{E}_{wji}w_i(z)$. The transfer function $G_{ji}(z)$ may in general be a cascade combination of subsystems $\{G_{ji\ell}(z)\}$:

$$G_{ji}(z) = \prod_{\ell} G_{ji\ell}(z) \tag{6}$$

The subsystems $G_{ji\ell}(z)$ may, for example, be a transfer function of a controller, an actuator, a plant, or a sensor associated with a position control system, process control system, magnetic levitation system, or other systems [4].

Expressing the frequency-domain model (5) in a linear regression form yields

$$y_{ji}(k) = \boldsymbol{\psi}_{ji}^T(k)\boldsymbol{\theta}_{ji} + v_{ji}(k) \tag{7}$$

where $v_{ij}(z) = D_j(z)\vartheta_{ij}(z)$; $\boldsymbol{\psi}_{ji}^T(k)$ is $1x2n_j$ regression vector formed of the regression vectors, formed $\boldsymbol{\psi}_{yji}^T(k)$ associated with $y_{ji}(k)$, and $\boldsymbol{\psi}_{ri}^T(k)$ associated with input $r_i(k)$:

$$\boldsymbol{\psi}_{ij}^T(k) = [\,\boldsymbol{\psi}_{yji}^T(k) \quad \boldsymbol{\psi}_{ri}^T(k)\,] \tag{8}$$

$\boldsymbol{\psi}_{yji}^T(k) = [\,-y_{ji}(k-1) \quad -y_{ji}(k-2) \quad . \quad -y_{ji}(k-n_j)\,]$; $\boldsymbol{\psi}_{ri}^T(k) = [\,r_i(k-1) \quad r_i(k-2) \quad . \; r_i(k-n_j)]$; $\boldsymbol{\theta}_{ji}$ is $2n_j$x1 feature vector formed of the n_j coefficients of the denominator polynomial $D_j(z)$ and the numerator polynomial $\boldsymbol{N}_{ij}(z)$:

$$\boldsymbol{\theta}_{ji} = [\,\boldsymbol{\theta}_{yj} \quad \boldsymbol{\theta}_{rji}\,]^T \tag{9}$$

Remarks: In the operational stage, we may not have access to the output $y_j(k)$, termed $y_{ji}(k)$, generated by the input $r_i(k)$ alone when rest of the inputs are set to zero. It is estimated during the identification phase of the multi-input and single-output model relating the accessible output $y_j(k)$ generated by all the inputs $\boldsymbol{r}(k)$.

2.2. Multi-input and single-output pairing

Using Eq. (5), the output $y_j(z)$ is the output due to all the inputs $\boldsymbol{r}(k)$ of MISO system, which is

$$y_j(z) = \sum_{i=1}^{p} y_{ji}(z) = \boldsymbol{G}_j(z)\boldsymbol{r}(z) + \vartheta_j(z) \tag{10}$$

where $\boldsymbol{G}_j(z) = D_j^{-1}(z)\boldsymbol{N}_j(z) = [\,G_{j1}(z) \quad G_{j2}(z) \quad . \quad G_{jp}(z)\,]$; $v_j(k) = \sum_{i=1}^{p} v_{ij}(k)$.

Expressing the frequency-domain model (10) in a linear regression form yields

$$y_j(k) = \boldsymbol{\psi}_j^T(k)\boldsymbol{\theta}_j + v_j(k)\,; j = 1, 2, 3, \ldots, q \tag{11}$$

where $\boldsymbol{\psi}_j^T(k)$ is $1x(n_j + n_jp)$ regression vector formed of the regression vectors $\boldsymbol{\psi}_{yj}^T(k)$ associated with $y_j(k)$, and $\boldsymbol{\psi}_r^T(k)$ associated with $\boldsymbol{r}(k)$:

$$\boldsymbol{\psi}_j^T(k) = [\,\boldsymbol{\psi}_{yj}^T(k) \quad \boldsymbol{\psi}_r^T(k)\,] \tag{12}$$

$\boldsymbol{\psi}_{yj}^T(k) = [\,-y_j(k-1) \quad -y_j(k-2) \quad . \quad -y_j(k-n_j)\,]$; $\boldsymbol{\psi}_r^T(k) = [\,\boldsymbol{\psi}_{r1}^T(k) \quad \boldsymbol{\psi}_{r2}^T(k) \quad . \quad \boldsymbol{\psi}_{rp}^T(k)\,]$; $\boldsymbol{\theta}_j$ is $(n_j + n_jp)$x1 feature vector formed of the n coefficients of the denominator polynomial $D_j(z)$ and the n_jp coefficients of the numerator polynomial $\boldsymbol{N}_j(z)$;

$$\boldsymbol{\theta}_j = [\,\boldsymbol{\theta}_{yj} \quad \boldsymbol{\theta}_{rj}\,]^T \tag{13}$$

where $\boldsymbol{\theta}_{yj} = [\,a_{j1} \quad a_{j2} \quad . \quad a_{jn_j}\,]^T$; $\boldsymbol{\theta}_{rj} = [\,\boldsymbol{\theta}_{rj1}^T \quad \boldsymbol{\theta}_{rj2}^T \quad . \quad \boldsymbol{\theta}_{rjp}^T\,]^T$.

2.3. Multi-input and multiple-output system

Extending the results of the time-domain expression to the MIMO (3), we get

$$\boldsymbol{y}(k) = \boldsymbol{\psi}^T(k)\boldsymbol{\theta} + \boldsymbol{v}(k) \tag{14}$$

where $\boldsymbol{\psi}^T(k)$ is $qx(n + npq)$ regression matrix formed of the regression vectors $\{\boldsymbol{\psi}_{ij}^T(k)\}$, and $\boldsymbol{\theta}$ is $(n + npq)$x1 feature vector formed of $\boldsymbol{\theta}_j$, $j = 1, 2, \ldots, q$ is given as follows:

$$\boldsymbol{\psi}^T(k) = \begin{bmatrix} \boldsymbol{\psi}_{y1}^T(k) & \boldsymbol{\psi}_r^T(k) & \mathbf{0} & \mathbf{0} & . & \mathbf{0} \\ \boldsymbol{\psi}_{y2}^T(k) & \mathbf{0} & \boldsymbol{\psi}_r^T(k) & \mathbf{0} & . & \mathbf{0} \\ \boldsymbol{\psi}_{y3}^T(k) & \mathbf{0} & \mathbf{0} & \boldsymbol{\psi}_r^T(k) & . & \mathbf{0} \\ . & . & . & . & . & . \\ \boldsymbol{\psi}_{yq}^T(k) & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & \boldsymbol{\psi}_r^T(k) \end{bmatrix}; \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_y \\ \boldsymbol{\theta}_{r1} \\ \boldsymbol{\theta}_{r2} \\ . \\ \boldsymbol{\theta}_{rp} \end{bmatrix} \tag{15}$$

The regression model (14) is the time-domain version of the frequency-domain model (3). Expressing the time-domain model (14) in the frequency domain, we get

$$\boldsymbol{y}(z) = \boldsymbol{\psi}^T(z)\boldsymbol{\theta} + \boldsymbol{v}(z) \tag{16}$$

2.4. Interconnected system

The system is an interconnection of subsystems such as the plant, the actuator, the sensors, and the controllers shown in **Figure 1**. Subfigure A at the top shows that jth output of the system $y_j = \sum_{i=1}^{p} y_{ji}$ is given by Eq. (10) where $y_{ij}(z)$ given in Eq. (5) is the output generated by the input r_i acting alone.

Subfigure B at the bottom shows that the transfer function $G_{ji}(z)$ in the path from the input r_i to the output y_{ij} is formed of subsystems $\{G_{ijl}(z)\}$. The subsystem $G_{ijl}(z)$ is driven by the input $u_{jil}(z)$ and its output is corrupted by the disturbance $w_{jil}(z)$. The input and the output of $G_{ji}(z)$ are r_i and y_{ji}, respectively, v_{ji} is the measurement noise, ϑ_{ji} given in Eq. (5) is the combined effect of the disturbances $\{w_{jik}\}$ and $\{v_{ji}\}$ on the output $y_{ji}(z)$.

2.5. Modeling of faults

There are two types of fault models, namely the additive and the multiplicative (or parametric) types. In the additive type, a fault is modeled as an additive exogenous input to the system, whereas in the multiplicative type, a fault is modeled as a change in the parameters, which completely characterize the fault behavior of the subsystems. Although the multiplicative and

http://dx.doi.org/10.5772/67870

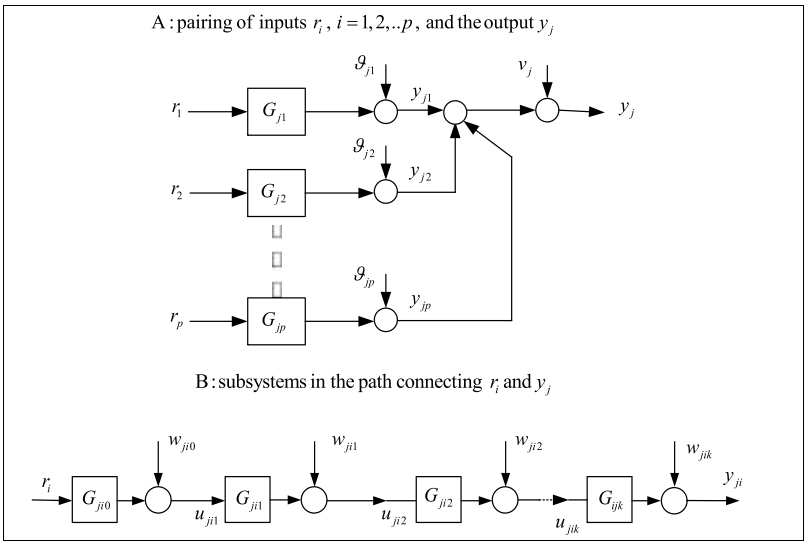


Figure 1. Pairing of the inputs and an output and the subsystem in the path *ji*.

additive perturbation models are equivalent, the multiplicative-type perturbation model is preferable. The multiplicative perturbation model of the cascade combination of subsystems can actually model the particular perturbation in any one of the subsystems under consideration.

3. Emulators

The emulator-based identification scheme is motivated by the model-free artificial neural network approach to capture the static and the dynamic behavior by presenting neural network data covering likely operating scenarios. An identified model at each operating point characterizes the behavior of the system in the neighborhood of that point. In practice, however, the system model may be perturbed because of variations in the parameters of the system. To overcome this problem, the system model is identified by performing a number of emulator parameter-perturbed experiments proposed in [4–5]. Each experiment consists of perturbing one or more emulator parameters. A linear model, termed *optimal model*, is identified as a best fit to the input-output data from the set of emulated perturbations. The optimal model thus obtained characterizes the behavior of the system over wider operating regions (in the neighborhood of the operating point), whereas the conventional model characterizes the behavior merely at the nominal operating point (i.e., the conventional approach assumes that the model of the system remains unperturbed at every operating point). The optimal model is more robust, that is, the identification errors resulting from the variations in the emulator parameters are significantly lower compared to those of the conventional one based on performing a single experiment (i.e., without using emulators).

During the system identification phase, a number of experiments are performed by (a) not perturbing the emulator parameters and (b) perturbing the emulator parameters one at a time, simultaneously perturbing two at a time, three at time, and so on till perturbing all of them. The input-output data collected from all experiments are termed emulator-generated data.

- *Nominal system model and the Kalman filter*: The emulator-generated data are used to identify the nominal optimal model of the system and the optimal Kalman filter model using the prediction error method.
- *Estimation of the influence vectors*: Using the least-squares method, the influence vectors are identified recursively using the input-output data obtained from the emulator-perturbed parameter experiments. First, the influence vector for influence vector for the single parameter perturbation is identified, and then using the estimated influence vector, the influence vector for the two simultaneous emulator perturbations is estimated. Generalizing, the influence vector for m simultaneous perturbation is identified, and then using all previous m estimates of the influence vectors, the $(m+1)$th influence vector is identified.

The emulators are transfer functions, which are connected in cascade with the subsystems to generate likely operating scenarios including normal and faulty one for reliable and accurate identification of the system, its associated Kalman filter, and the feature vector-emulator map.

Emulators are connected to the system during the identification phase and its parameter is varied to generate likely operating scenarios. During the operational phase, the static emulators are disconnected, as it were, by setting them to unit values. The dynamic emulator, however, is not disconnected. Its gain is set to unity and its phase made a non-zero negligibly small value so that (a) both of these parameters have a negligible effect on the dynamic behavior of the system during the operational phase and (b) the order of the system during the identification and the operational phases remains identical to ensure mathematical tractability without causing performance degradation. The role of the emulator-generated data includes the following:

3.1. Emulator-generated data for MISO system

The MISO system is given by Eq. (11) relating all the inputs $\boldsymbol{r}(k)$ and the output $y_j(k)$ identified by connecting an emulator $E_j(z)$ in cascade with $\boldsymbol{r}(z)$. The emulator is a first-order all-pass filter given by

$$E_j(z) = \gamma_{j2}\left(\frac{\gamma_{j1} + z^{-1}}{1 + \gamma_{j1}z^{-1}}\right) \tag{17}$$

where $|\gamma_{j1}| < 1$ to ensure stability. The emulators γ_{j1} and γ_{j2} are varied one at a time, and both simultaneously. During the identification, an emulator $E_j(z)$, which is a first-order all-pass filter (17), is connected to the input $r_j(k)$ in cascade with nominal model $\boldsymbol{G}_{j0}(z)$. A number of experiments are performed by varying the emulator parameters γ_{j1}, γ_{j2} one at a time and both simultaneously to acquire emulator-generated data: it is assumed for simplicity that the same

input is applied to all the experiments. Using Eq. (10), the MISO model relating $r_j(k)$ and $y_j(k)$ becomes

$$y_j^{el}(z) = \boldsymbol{G}_{j0}(z)E_j(z)\boldsymbol{r}(z) + \vartheta_j^e(z), \quad el = 1, 2, \ldots, n_{\exp}, \quad l = 1, 2, 3 \tag{18}$$

where $y_j^{e1}(z)$, $y_j^{e2}(z)$, and $y_j^{e3}(z)$ denote, respectively, the output generated by varying γ_{j1}, γ_{j2} and both γ_{j1}, γ_{j2}.

3.2. Emulator-generated data for SISO system

The feature vector-emulator map of the SISO system (5) is estimated for the isolation of faults in the subsystems $\{G_{ij\ell}(z)\}$. The emulators $E_{ji\ell}(z)$ are connected to an accessible input or output $\{u_{ji\ell}\}$ in cascade with the subsystems $\{G_{ji\ell}(z)\}$ to mimic their variations. In other words, the known emulator parameter variations mimic those of the unknown parameters of the associated subsystems. The accessible inputs include the tracking error, the control input, actuator input, and sensor output.

The emulator $E_{ji\ell}(z)$ may be a dynamic system, a constant gain $(\gamma_{ji\ell})$, a gain, and a pure delay of d time instants $(\gamma_{ji\ell}z^{-d})$, a first-order all-pass filter $\left(\frac{\gamma_{ji\ell}+z^{-1}}{1+\gamma_{ji\ell}z^{-1}}\right)$ or a Blaschke product of all first-order-pass filters $\left(\prod_{\ell}\frac{\gamma_{ji\ell}+z^{-1}}{1+\gamma_{ji\ell}z^{-1}}\right)$ [3]. The emulator $E_{ji}(z)$ is chosen to be a product of a static gain and a first-order all-pass filter to mimic the behavior of the subsystem $G_{ji}(z) = \prod_{\ell=1}^{l} G_{ji\ell}(z)$ of the SISO system given by Eqs. (5) and (6)

$$\gamma_{ji2}\left(\frac{\gamma_{ji1}+z^{-1}}{1+\gamma_{ji1}z^{-1}}\right) \tag{19}$$

In order to ensure stability of the dynamic emulator, parameter γ_{ji1} is constrained by $|\gamma_{ji\ell}| < 1$.

Connecting the emulator $E_{ji}(z)$ given in Eq. (19) to the nominal SISO model $G_{ji0}(z)$ using Eqs. (5) and (6), we get

$$y_{ji}^{el}(z) = E_{ji}(z)G_{ji0}(z)r_i(z) + \vartheta_{ji}^e(z), \quad e = 1, 2, \ldots, n_{\exp}, \quad l = 1, 2, 3 \tag{20}$$

where $y_{ji}^{e1}(z)$, $y_{ji}^{e2}(z)$, and $y_{ji}^{e3}(z)$ denote, respectively, the output generated by varying γ_{ji1}, γ_{ji2} and both γ_{ji1} and γ_{ji2}.

Figure 2 shows an example of a closed-loop position control system formed of a controller, an actuator, a plant, and a sensor in the path connecting the tracking error $e_{ri}(k) = r_i(k) - y_{ji}(k)$ and the output y_{ji}. Only $e_{ri}(k)$, $u_{ij1}(k)$, and $u_{ij3}(k)$ are the measurement outputs. The emulators $E_{ji1}(z) = \frac{\gamma_{ji1}+z^{-1}}{1+\gamma_{ji1}z^{-1}}$, and $E_{ji2} = \gamma_{ji2}$ are connected to u_{ji1}, and $E_{ji3} = \gamma_{ji3}$ is connected to u_{ji3} to mimic

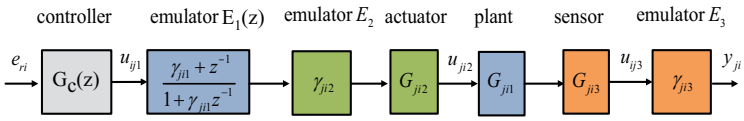


Figure 2. Position control system: emulators and subsystems.

the perturbations in the dynamic plant $G_{ji1}(z)$, the static actuator $G_{ji2}(z) = k_A$ and the static sensor $G_{ji3}(z) = k_s$, respectively, where $E_{ji1}(z)$ is dynamic, and E_{ji2} and E_{ji3} are static emulators.

The nominal static emulator is set to unit value $\gamma_{ijk}^0 = 1$. The variation $\Delta\gamma_{jik}$ of an emulator γ_{ijk} may be expressed in terms of its nominal value γ_{jik}^0 as $\Delta\gamma_{jik} = \gamma_{jik} - \gamma_{jik}^0$.

3.3. Feature vector-emulator map

The feature vector-emulator map for the SISO and the MISO systems is developed subsequently.

3.3.1. SISO system

Consider the emulator-perturbed SISO system (20) relating the inputs $r_i(k)$ and $y_{ji}(k)$ and the associated linear regression model (7). The feature vector $\boldsymbol{\theta}_{ji}$ is a nonlinear function of the emulator parameter $\boldsymbol{\gamma}^{ji} = [\,\gamma_{ji1} \quad \gamma_{ji2}\,]$. Assuming that the feature vector $\boldsymbol{\theta}_{ji}$ is a continuous function of $\boldsymbol{\gamma}^{ji}$, then using Weierstrass approximation theorem, the feature vector-emulator map becomes

$$\Delta\boldsymbol{\theta}_{ji} = \boldsymbol{\Omega}_{ji1}\Delta\gamma_{ji1} + \boldsymbol{\Omega}_{ji2}\Delta\gamma_{ji2} + \boldsymbol{\Omega}_{ji12}\Delta\gamma_{ji1}\Delta\gamma_{ji2} \tag{21}$$

where $\Delta\boldsymbol{\theta}_{ji} = \boldsymbol{\theta}_{ji} - \boldsymbol{\theta}_{ji}^0; \Delta\gamma_{j\ell} = \gamma_{j\ell} - \gamma_{j\ell}^0$ is the parameter variation; $\boldsymbol{\theta}_{ji}^0$ is the nominal feature vector; $\boldsymbol{\Omega}_{ji1}$ is a $2n_j x1$ vector of partial derivatives of the feature vector $\boldsymbol{\theta}_{ji}$ with respect to γ_{ji1} evaluated at the unperturbed nominal emulator value γ_{ji1}^0. Similarly, $\boldsymbol{\Omega}_{ji2}$ is a $2n_j x1$ vector of partial derivatives of the feature vector $\boldsymbol{\theta}_{ji}$ with respect to γ_{ji2} evaluated at the unperturbed nominal emulator value γ_{ji2}^0, $\boldsymbol{\Omega}_{ji12}$ is the second partial derivatives with respect to γ_{ji1} and γ_{ji2} evaluated at the unperturbed nominal emulator value γ_{ji1}^0 and γ_{ji2}^0. The partial derivative terms $\boldsymbol{\Omega}_{ji1}$, $\boldsymbol{\Omega}_{ji2}$ $\boldsymbol{\Omega}_{ji12}$, which are the Jacobean of the feature vector $\boldsymbol{\theta}_{ji}$ with respect to the emulator parameters $\{\gamma_{jik}\}$, are termed *influence vectors*. The influence vectors play a crucial role in isolating a fault occurring in any subsystem. The influence vectors $\boldsymbol{\Omega}_{ji1}$, $\boldsymbol{\Omega}_{ji2,}$ and $\boldsymbol{\Omega}_{ji12}$ track the degree of variations in the parameters of the subsystem perturbations.

Substituting for $\boldsymbol{\theta}_{ji}$ in (7), the variation $\Delta y_{ji}(k) = y_{ji}(k) - y_{ji}^0(k)$ between the actual output $y_{ji}(k)$ and the nominal fault-free output $y_{ji}^0(k)$ becomes

$$\Delta y_{ji}(k) = \boldsymbol{\psi}_{ji}^T(k)(\boldsymbol{\Omega}_{ji1}\Delta\gamma_{ji1} + \boldsymbol{\Omega}_{ji2}\Delta\gamma_{ji2} + \boldsymbol{\Omega}_{ji12}\Delta\gamma_{ji1}\Delta\gamma_{ji2}) + \upsilon_{ji}(k) \tag{22}$$

Let $\boldsymbol{\Omega}_{ji}$ be an influence matrix associated with the emulators located at the path ij

$$\boldsymbol{\Omega}_{ji} = [\,\boldsymbol{\Omega}_{jik} \quad \boldsymbol{\Omega}_{jik\ell} \quad \boldsymbol{\Omega}_{jik\ell mn} \quad . \quad \boldsymbol{\Omega}_{ji12\ldots q}\,] \tag{23}$$

A number of emulator parameter-perturbed experiments are performed by perturbing the parameters of the emulators (20). For each experiment, N input-output data $(\boldsymbol{y}_j^e(k), \boldsymbol{r}(k))$ are obtained, $k = 1, 2, \ldots, N$. The input $\boldsymbol{r}(k)$ for each experiment is chosen to be persistently exciting. The regression models associated with the experiments and Eq. (22) are given as follows:

$$\begin{aligned}
&\Delta y_{ji}^{e1}(k) = \boldsymbol{\psi}_{ji}^T(k)\Delta\gamma_{ji1}\,\boldsymbol{\Omega}_{ji1} + \boldsymbol{\upsilon}_{ji}^{e1}(k)\\
&\Lambda y_{ji}^{e2}(k) - \boldsymbol{\psi}_{ji}^T(k)\Delta\gamma_{ji1}\,\boldsymbol{\Omega}_{ji1} = \boldsymbol{\psi}_{ji}^T(k)\boldsymbol{\Omega}_{ji2}\Delta\gamma_{ji2} + \upsilon_{ji}^{e2}(k)\\
&\Delta y_{ji}^{e3}(k) - \boldsymbol{\psi}_{ji}^T(k)(\Delta\gamma_{ji1}\,\boldsymbol{\Omega}_{ji1} + \boldsymbol{\Omega}_{ji2}\Delta\gamma_{ji2}) = \boldsymbol{\psi}_{ji}^T(k)(\boldsymbol{\Omega}_{ji12}\Delta\gamma_{ji12}) + \upsilon_{ji}^{e3}(k)
\end{aligned} \tag{24}$$

3.3.2. MISO system

Consider the emulator-perturbed MISO system (18) relating the inputs $\boldsymbol{r}(k)$ and $y_j(k)$, and the associated linear regression model (11). Similar to Eqs. (21) and (24), we get

$$\Delta\boldsymbol{\theta}_j = \boldsymbol{\Omega}_{j1}\Delta\gamma_{j1} + \boldsymbol{\Omega}_{j2}\Delta\gamma_{j2} + \boldsymbol{\Omega}_{j12}\Delta\gamma_{j1}\Delta\gamma_{j2} \tag{25}$$

$$\begin{aligned}
&\Delta y_j^{e1}(k) = \boldsymbol{\psi}_j^T(k)\Delta\gamma_{j1}\,\boldsymbol{\Omega}_{j1} + \boldsymbol{\upsilon}_j^{e1}(k)\\
&\Lambda y_j^{e2}(k) - \boldsymbol{\psi}_j^T(k)\Delta\gamma_{j1}\,\boldsymbol{\Omega}_{j1} = \boldsymbol{\psi}_j^T(k)\boldsymbol{\Omega}_{j2}\Delta\gamma_{j2} + \upsilon_j^{e2}(k)\\
&\Delta y_j^{e3}(k) - \boldsymbol{\psi}_j^T(k)(\Delta\gamma_{j1}\,\boldsymbol{\Omega}_{j1} + \boldsymbol{\Omega}_{j2}\Delta\gamma_{j2}) = \boldsymbol{\psi}_j^T(k)(\boldsymbol{\Omega}_{j12}\Delta\gamma_{j12}) + \upsilon_j^{e3}(k)
\end{aligned} \tag{26}$$

4. Identification

The prediction error method can be derived from the residual model of the Kalman filter, which is presented in the next section. It is used to identify both the nominal system and the Kalman filter associated with the system without the need for a priori knowledge of the covariances of the noise and the disturbance statistics. Prediction error method is consistent, efficient, and a gold standard for system identification, and can identify open-loop and closed-loop systems. The variance the parameter estimates asymptotically approaches the Cramer-Rao lower bound.

Optimal models: The optimal system and the associated Kalman filter are identified using the prediction error method using computationally efficient scheme. First, the MISO system is identified and then the SISO system is derived from the estimate of feature vector associated with the MISO system. The emulator-generated data generated using Eq. (18) are used to identify MISO system (10) and the nominal feature vector $\boldsymbol{\theta}_j^0$ for Eq. (11), which is the best least-squared fit to set all perturbed feature vector $\boldsymbol{\theta}_j$, and the Kalman gain $\boldsymbol{K}_{j0}$ are estimated.

Let the optimal state-space model of the MISO system be $(\boldsymbol{A}_{j0}, \boldsymbol{B}_{j0}, \boldsymbol{C}_{j0})$ and associated Kalman filter be $(\boldsymbol{A}_{j0} - \boldsymbol{K}_{j0}\boldsymbol{C}_{j0}, [\boldsymbol{K}_{j0}\ \boldsymbol{B}_{j0}], \boldsymbol{C}_{j0})$. Let the optimal transfer matrix of the MISO system and the optimal estimate of the output be $\boldsymbol{G}_j^{opt}(z)$ and $\hat{y}_j^{opt}(k)$, respectively. Using Eq. (10), we get

$$\hat{y}_j^{opt}(z) = \boldsymbol{G}_j^{opt}(z)\boldsymbol{r}(z) + \vartheta_j(z) \tag{27}$$

Then, the best estimate of the feature vector $\boldsymbol{\theta}_{ji}$ of the SISO system (7), denoted $\boldsymbol{\theta}_{ji}^0$, and the Kalman gain are estimated from $\boldsymbol{\theta}_j^0$.

4.1. Estimation of the influence vectors

SISO system: Knowing the emulator parameter perturbations $\Delta\gamma_{ji1}$, $\Delta\gamma_{ji2}$, $\Delta\gamma_{ji12}$ and the resulting emulator-generated data, the influence vectors $\hat{\boldsymbol{\Omega}}_{ji1}$, $\hat{\boldsymbol{\Omega}}_{ji2}$, and $\hat{\boldsymbol{\Omega}}_{ji12}$ are estimated recursively using the least-squares method using Eq. (24)

$$\begin{aligned}
\hat{\boldsymbol{\Omega}}_{ji1} &= \arg\min_{\boldsymbol{\Omega}_{ji1}} \left\{ \left\| \Delta y_{ji}^{ei}(k) - \boldsymbol{\psi}_{ji}^T(k)\boldsymbol{\Omega}_{ji1}\Delta\gamma_{ji1} \right\|^2 \right\} \\
\hat{\boldsymbol{\Omega}}_{ji2} &= \arg\min_{\boldsymbol{\Omega}_{ji2}} \left\{ \left\| \Lambda y_{ji}^{e2}(k) - \boldsymbol{\psi}_{ji}^T(k)\Delta\gamma_{ji1}\ \hat{\boldsymbol{\Omega}}_{ji1} - \boldsymbol{\psi}_{ji}^T(k)\boldsymbol{\Omega}_{ji2}\Delta\gamma_{ji2} \right\|^2 \right\} \\
\hat{\boldsymbol{\Omega}}_{ji12} &= \arg\min_{\boldsymbol{\Omega}_{jiklm}} \left\{ \left\| \Delta y_j^{e3}(k) - \boldsymbol{\psi}_{ji}^T(k)(\Delta\gamma_{ji1}\ \hat{\boldsymbol{\Omega}}_{ji1} + \hat{\boldsymbol{\Omega}}_{ji2}\Delta\gamma_{ji2}) - \boldsymbol{\psi}_{ji}^T(k)\boldsymbol{\Omega}_{ji12}\Delta\gamma_{ji12} \right\|^2 \right\}
\end{aligned} \tag{28}$$

where $\|x(k)\|^2 = \sum_{k=1}^{N} x^2(k)$.

MISO system: Similar to Eq. (28), the influence vectors $\hat{\boldsymbol{\Omega}}_{jk}$, $\hat{\boldsymbol{\Omega}}_{j2}$, and $\hat{\boldsymbol{\Omega}}_{j12}$ are estimated.

5. Model of the Kalman filter

The Kalman filter forms the backbone of the MISO and the SISO systems fault detection and for fault isolation, respectively. The Kalman filter is a closed-loop system, which is (a) an exact copy of the identified nominal of the system driven by the residual, which is the error between the output and its estimate, and (b) is stabilized by the Kalman gain.

MISO system: Using the state-space model $(\boldsymbol{A}_{j0}, \boldsymbol{B}_{j0}, \boldsymbol{C}_{j0})$ derived from the identified nominal feature vector $\boldsymbol{\theta}_j^0$. The Kalman filter $(\boldsymbol{A}_{j0} - \boldsymbol{K}_{j0}\boldsymbol{C}_{j0}, [\boldsymbol{K}_{j0}\ \boldsymbol{B}_{j0}], \boldsymbol{C}_{j0})$ associated with the MISO system (10) is

$$\begin{aligned}
\hat{\boldsymbol{x}}_j(k+1) &= (\boldsymbol{A}_{j0} - \boldsymbol{K}_{j0}\boldsymbol{C}_{j0})\hat{\boldsymbol{x}}_j(k) + \boldsymbol{K}_{j0}y_j(k) + \boldsymbol{B}_{j0}\,\boldsymbol{r}(k) \\
\hat{y}_j(k) &= \boldsymbol{C}_{j0}\hat{\boldsymbol{x}}_j(k) \\
e_j(k) &= y_j(k) - \hat{y}_j(k)
\end{aligned} \tag{29}$$

where $\hat{\boldsymbol{x}}_j(k)$ and $\hat{y}_j(k)$ are, respectively, the minimum variance estimates of the state and the output.

Figure 3 shows the nominal fault-free system and the Kalman filter. The structure of the Kalman filter is based on the internal model principle, which embodies the nominal system model $(\boldsymbol{A}_{j0}, \boldsymbol{B}_{j0}, \boldsymbol{C}_{j0})$. The inputs to the Kalman filter are the input $\boldsymbol{r}(k)$ and the output $y_j(k)$ which is corrupted by the disturbance $w_j(k)$ and the measurement noise $v_j(k)$.

5.1. Expressions of the residual

The expression for the residuals for the MISO system $e_j(z)$ and the SISO system $e_{ji}(z)$ is derived from the Kalman filter (29).

MISO model: The frequency-domain expression, relating the $n_u x1$ input $\boldsymbol{r}(z)$ and output $y_j(z)$ to the residual $e_j(z)$ is given by the following model, termed *residual model*:

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)} y_j(z) - \frac{\overline{\boldsymbol{N}}_{j0}(z)}{F_{j0}(z)} \boldsymbol{r}(z) \tag{30}$$

where $F_{j0}(z) = |z\boldsymbol{I} - \boldsymbol{A}_{j0} + \boldsymbol{K}_{j0}\boldsymbol{C}_{j0}|$ is the characteristic polynomial termed *Kalman polynomial;*
$\overline{D}_{j0}(z) = F_{j0}(z)\left(\boldsymbol{I} - \boldsymbol{C}_{j0}(\boldsymbol{zI} - \boldsymbol{A}_{j0} + \boldsymbol{K}_{j0}\boldsymbol{C}_{j0})^{-1}\boldsymbol{K}_{j0}\right)$ -
$\overline{\boldsymbol{N}}_{j0}(z) = [\,\overline{N}_{j10}(z) \quad \overline{N}_{j10}(z) \quad . \quad \overline{N}_{jp0}(z)\,] = F_{j0}(z)\left(\boldsymbol{C}_{j0}(z\boldsymbol{I} - \boldsymbol{A}_{j0} + \boldsymbol{K}_{j0}\boldsymbol{C}_0)^{-1}\boldsymbol{B}_{j0}\right)$

SISO system: The residual $e_{ji}(z)$ is derived from the residual model (30) from the map relating $e_{ji}(z)$ to $y_j(z)$ and $r_i(z)$:

$$e_{ji}(z) = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)} y_j(z) - \frac{\overline{N}_{ji0}(z)}{F_{j0}(z)} r_i(z) \tag{31}$$

where $\overline{N}_{0ji}(z)$ is the i^{th} element of $\overline{\boldsymbol{N}}_{j0}(z)$.

5.1.1. Key properties of the Kalman filter residual

The Kalman filter forms the backbone of the proposed scheme in view of its key properties proved in [1]. These properties exploited in developing the system identification using the

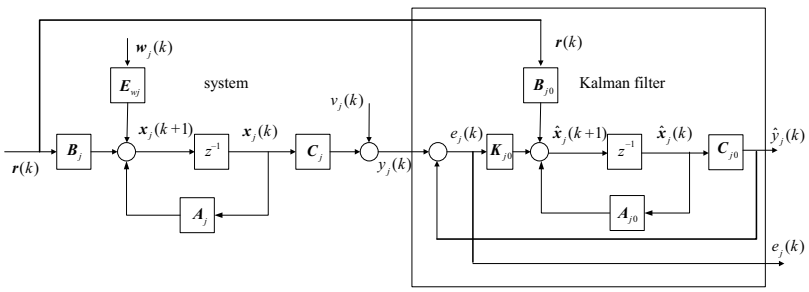


Figure 3. The system and its associated Kalman filter.

residual model, and in unified approach to fault detection and isolation fault, where a fault is defined as an incipient fault resulting in the model mismatch.

5.2. Propositions

We establish important results, in the form of lemmas that are crucial to the development of the proposed fault diagnosis scheme. In *Lemma 1,* it is shown that (a) the system transfer function can be estimated from the residual model and (b) Kalman filter whitens the output error $\vartheta_j(z)$ given in Eq. (10). *Lemma 2* shows that the residual will not be a zero-mean white noise process if there is a model mismatch, and there will be an additive fault indicating term, which is a function of the deviation between the actual feature vector $\boldsymbol{\theta}_j$ of the system model $(\boldsymbol{A}_j, \boldsymbol{B}_j, \boldsymbol{C}_j)$ and the nominal fault-free feature vector $\boldsymbol{\theta}_j^0$ of nominal fault-free model $(\boldsymbol{A}_{j0}, \boldsymbol{B}_{j0}, \boldsymbol{C}_{j0})$.

Case 1: The system and the nominal models are identical

Lemma 1:

$$\boldsymbol{G}_{j0}(z) = D_{j0}^{-1}(z)\boldsymbol{N}_{j0}(z) = \overline{D}_{j0}^{-1}(z)\overline{\boldsymbol{N}}_{j0}(z) \tag{32}$$

where $\boldsymbol{G}_{j0}(z)$ is the transfer function of the nominal fault-free model $(\boldsymbol{A}_{j0}, \boldsymbol{B}_{j0}, \boldsymbol{C}_{j0})$.

Proof: Substituting for $\boldsymbol{y}(z)$ from Eq. (10), the residual model (30) becomes

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\left(\left(D_j^{-1}(z)\boldsymbol{N}_j(z) - \overline{D}_{j0}^{-1}(z)\overline{\boldsymbol{N}}_{j0}(z)\right)\boldsymbol{r}(z)\right) + \frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\vartheta_j(z) \tag{33}$$

Correlating both sides with input $\boldsymbol{r}(k)$, and invoking the orthogonality properties, the residual, namely $\boldsymbol{r}(k)$, is uncorrelated with both $e_j(k)$ and the output error $\vartheta_j(k)$ [4], we get

$$\frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\left(\left(D_j^{-1}(z)\boldsymbol{N}_j(z) - \overline{D}_{j0}^{-1}(z)\overline{\boldsymbol{N}}_{j0}(z)\right)\boldsymbol{r}(z)\right) = 0 \tag{34}$$

Hence, Eq. (32) holds.

Corollary 1: The filter $\frac{\overline{D}_{j0}(z)}{F_{j0}(z)}$ whitens the output error $\vartheta_j(z)$ if there is no model mismatch:

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\vartheta_j(z) \tag{35}$$

Proof: Consider the expression for the model-matching case (33). Using Eq. (32), we establish Eq. (35).

Case 2: System and the nominal model mismatch

Lemma 2: If there is model mismatch, then

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\Delta G_j(z) + \vartheta_{jf} \tag{36}$$

$$e_j(z) = \boldsymbol{\psi}_{jf}^T(z)\Delta\boldsymbol{\theta}_j + v_{jf}(z) \tag{37}$$

where $\Delta G_j(z) = D_j^{-1}(z)\boldsymbol{N}_j(z) - D_{j0}^{-1}(z)\boldsymbol{N}_{j0}(z)$, $\Delta\boldsymbol{\theta}_j = \boldsymbol{\theta}_j - \boldsymbol{\theta}_j^0$; $\boldsymbol{\psi}_{jf}^T(z) = \frac{\overline{D}_{j0}(z)}{D_j(z)F_{j0}(z)}\boldsymbol{\psi}_j^T(z)$, $\vartheta_{jf} = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\vartheta_j(z)$, and $e_{jf}(z) = \frac{\overline{D}_{j0}(z)}{D_j(z)F_{j0}(z)}v_j(z)$ are the filtered regression matrix $\boldsymbol{\psi}_j^T(z)$ and filtered output error $\vartheta_j(z)$, filtered equation error $v_j(z)$, respectively.

Proof:

Case 1: Consider expression (33). Using Eq. (32), we get

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\left(\left(D_j^{-1}(z)\boldsymbol{N}_j(z) - D_{j0}^{-1}(z)\boldsymbol{N}_{j0}(z)\right)\boldsymbol{r}(z)\right) + \vartheta_{jf}(z) \tag{38}$$

Substituting $\Delta G_j(z) = D_j^{-1}(z)\boldsymbol{N}_j(z) - D_{j0}^{-1}(z)\boldsymbol{N}_{j0}(z)$, we get Eq. (36).

Case 2: Expressing the residual model (30) in an alternative form:

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{F_{j0}(z)}\left(y_j(z) - \overline{D}_{j0}^{-1}(z)\overline{\boldsymbol{N}}_{j0}(z)\boldsymbol{r}(z)\right) \tag{39}$$

Using Eq. (32) and re-arranging, we get

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{D_{j0}(z)F_{j0}(z)}\left(D_{j0}(z)y_j(z) - \overline{\boldsymbol{N}}_{j0}(z)\boldsymbol{r}(z)\right) \tag{40}$$

Adding and subtracting $y_j(z)$ inside the bracket on the right-hand side yields

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{D_{j0}(z)F_{j0}(z)}\left(y_j(z) - \left(1 - D_{j0}(z)\right)y_j(z) - \overline{\boldsymbol{N}}_{j0}(z)\boldsymbol{r}(z)\right) \tag{41}$$

Using the expression for the regression model (11) and substituting for the actual and the nominal fault-free cases, we get

$$e_j(z) = \frac{\overline{D}_{j0}(z)}{D_{j0}(z)F_{j0}(z)}\boldsymbol{\psi}_j^T(k)\Delta\boldsymbol{\theta}_j + v_{jf}(k) \tag{42}$$

Remarks: If there is a model mismatch because of variations in the subsystem parameters, the residual is no longer zero-mean white noise process. The residual has an additive term, which is affine in the deviation in the system transfer function $\Delta G_j(z)$ or equivalently affine in the feature vector $\boldsymbol{\psi}_{jf}^T(z)\Delta\boldsymbol{\theta}_j$. The additive terms are termed fault indicators. This shows that the Kalman filter provides a unifying approach to handle both fault detection and fault isolation.

In view of the key properties, the Kalman filter is employed for identification and the fault diagnosis. In system identification, the criterion for determining whether the identified model has captured completely the dynamic behavior of the system is that the residual (error between the output and its estimate obtained using the identified model) is a zero-mean white noise process. Consider the problem of identification of the system. Since the equation error $\boldsymbol{v}(k)$ is a colored noise process, the parameter estimates will be biased and inefficient. To overcome this, the input and the output are whitened using the Kalman filter as shown in Eq. (35) of Corollary 1. The Kalman filter model (29) may be interpreted as an inverse system generating the innovation sequence $\boldsymbol{e}(k)$, or alternatively as a whitening-filter realization of a state-space model that is driven by both the disturbance and measurement noise.

Lemma 3

$$e_{ji}(z) = \boldsymbol{\psi}_{jif}^{T}(z)\Delta\boldsymbol{\theta}_{ji} + v_{jif}(z) \tag{43}$$

where $F_{ji0}(z) = |z\boldsymbol{I} - \boldsymbol{A}_{ji0} + \boldsymbol{K}_{ji0}\boldsymbol{C}_{j0}|$, $\boldsymbol{\psi}_{jif}^{T}(z) = \frac{\overline{D}_{j0}(z)}{D_j(z)F_{ji0}(z)}\boldsymbol{\psi}_{ji}^{T}(z)$;$v_{jif}(z) = \frac{\overline{D}_{0j}(z)}{D_j(z)F_{ji0}(z)}v_{ji}(z)$

Proof: The proof follows from Eqs. (31) and (37).

6. Bayesian approach fault diagnosis

The objective of fault detection is to assert whether the given residual belongs to a set of fault-free data or faulty residual data, while fault isolation is determined to which class of emulator-perturbed residual the given data belong. The problem of fault detection and fault isolation is formulated by a pattern classification problem. Fault detection is a binary pattern classification, while the fault isolation is a multi-class pattern classification. The Bayesian decision strategy is employed to assert appropriate class label. The Bayesian decision strategy is based on the a posteriori conditional probability of deciding a hypothesis given the data, a priori probability of the hypothesis, and a performance measure. The decision strategy is determined from the minimization of the performance measure with respect to all hypotheses.

The $Nx1$ residual $\boldsymbol{e}(k)$ is located in a different region of the N-dimensional plane depending upon the fault type. In the ideal case regions, there will not be overlaps between regions associated with different fault types. However, due to noise, disturbances, and other measurement artifacts there will be overlap between the various regions. Hence, Bayesian strategy is employed to asset an appropriate class label to ensure a high-probability correct decision, and a low probability of false alarms.

6.1. Fault detection

Fault detection is posed as a binary hypothesis-testing problem. The criterion to choose between the two hypotheses, namely the presence or an absence of a fault, is based on minimizing the Bayes risk, which quantifies the costs associated with correct and incorrect decisions. The $Nx1$ Kalman filter residual data $\boldsymbol{e}(k)$ generated by Eq. (29) is employed. The

minimization of the Bayes risk yields the likelihood ratio test. The decision between the two hypotheses is based on comparing the likelihood ratio, which is the ratio of the conditional probabilities under the two hypotheses, to a threshold value. The resulting binary composite hypothesis-testing problem compares the test statistics of residual $e(k)$ with a threshold value η:

$$t_s(e) \begin{cases} \leq \eta & \text{no fault} \\ > \eta & \text{fault} \end{cases} \tag{44}$$

The test statistics depends upon the input $r(k)$ that generates the residual $e(k)$ [4]:

$$t_s(e) = \begin{cases} \left| \frac{1}{N} \sum_{i=k-N+1}^{k} e(i) \right| & r(k) = \text{constant} \\ P_{ee}(f_0) & r(k) \text{ is a sinusoid} \\ \frac{1}{N} \sum_{i=k-N+1}^{k} e^2(i) & r(k) \text{ is an arbitrary signal} \end{cases} \tag{45}$$

6.1.1. Computationally efficient scheme

A computationally efficient scheme is employed here for the detection:

- The status of each of the MISO systems $\boldsymbol{G}_j(z)$ relating all the inputs $\boldsymbol{r}(z)$ and all the outputs $y_j(z)$ is evaluated for all $j = 1, 2, \ldots, q$ using the binary hypothesis scheme (44). Using the test statistics of the residuals $e_j(k)$ given by Eq. (30) yields

$$t_s(e_j) \begin{cases} \leq \eta_j & \text{no fault} \\ > \eta_j & \text{fault} \end{cases}, \quad j = 1, 2, 3, \ldots, q \tag{46}$$

- If a fault is asserted in $\boldsymbol{G}_j(z)$, then the status of each of the p subsystems $G_{ji}(z)$ of the SISO system is asserted using the test statistics of the residuals $e_{ji}(k)$ (31):

$$t_s(e_{ji}) \begin{cases} \leq \eta_{ji} & \text{no fault} \\ > \eta_{ji} & \text{fault} \end{cases}, \quad i = 1, 2, 3, \ldots, p \tag{47}$$

Fault accommodation: If a fault is asserted, then the Kalman gain is adapted online, the system re-identified, and the Kalman filter redesigned accordingly, thus the fault is accommodated and, in the extreme case, the system is shut down for safety reasons.

7. Evaluation on simulated system

The proposed emulator-based system identification of the system, the associated Kalman filter, feature vector-emulator map, and finally the fault diagnosis are illustrated using an example of a position control system formed of an actuator, a sensor, and a plant.

7.1. System model

A two-input and two-output system fault-free system $(\boldsymbol{A}_0, \boldsymbol{B}_0, \boldsymbol{C}_0)$ given by Eq. (1) is considered where

$$\boldsymbol{A}_0 = \begin{bmatrix} 0 & -0.7 & 0 & 0 \\ 1 & 1.5 & 0 & 0 \\ 0 & 0 & 0 & -0.82 \\ 0 & 0 & 1 & 1.8 \end{bmatrix}; \quad \boldsymbol{B}_0 = \begin{bmatrix} 0.5 & 1 \\ 1 & 0 \\ 1 & -0.3 \\ 0 & 1 \end{bmatrix}; \quad \boldsymbol{C}_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{48}$$

The nominal transfer matrix of the MIMO system (3) is

$$\boldsymbol{G}_0(z) = \begin{bmatrix} G_{11}(z) & G_{12}(z) \\ G_{21}(z) & G_{22}(z) \end{bmatrix} = \begin{bmatrix} \dfrac{1+z^{-1}}{1-1.5z^{-1}+0.7z^{-2}} & \dfrac{1}{1-1.5z^{-1}+0.7z^{-2}} \\ \dfrac{1}{1-1.8z^{-1}+0.82z^{-2}} & \dfrac{1-0.2z^{-1}}{1-1.8z^{-1}+0.82z^{-2}} \end{bmatrix} \tag{49}$$

The nominal MISO transfer matrix, $\boldsymbol{G}_{j0}(z) = D_0^{-1}(z)\boldsymbol{N}_{j0}(z)$, $j = 1, 2$, of the system is

$$\begin{aligned} D_0(z) &= 1 - 3.3z^{-1} + 4.22z^{-2} - 2.49z^{-3} + 0.574z^{-4} \\ \boldsymbol{N}_0(z) &= \begin{bmatrix} z^{-1} - 1.3z^{-2} & -0.08z^{-1} + 0.41z^{-2} \\ z^{-1} - 1.8z^{-2} & 1.15z^{-1} - 0.2z^{-1} \end{bmatrix} \end{aligned} \tag{50}$$

Figure 4a shows the emulator-generated MISO output 1, y_1^{e1} MISO output 2, y_2^{e1}, SISO output 11, y_{11}^{e1}, SISO output 12, y_{12}^{e1}, SISO output 21, y_{21}^{e1} and SISO output 22, y_{22}^{e1} given in Eqs. (18) and (20) resulting from the variations of the emulator parameters γ_{j1} and γ_{ji1}, respectively. Subfigures A and B show plots of the perturbed step responses $y_1^{e1}(k)$ and $y_2^{e1}(k)$ with respect to time, while subfigures C–F show plots of the perturbed outputs $y_{11}^{e1}(k)$, $y_{12}^{e1}(k)$, $y_{21}^{e1}(k)$, and $y_{22}^{e1}(k)$ with respect to time. The outputs are in centimeters (cm) and the time is in seconds (s). The plots are generated when the emulator parameter γ_{j1} is varied. The variations $\Delta\gamma_{j1}$ are $\{0.1 \quad 0.5 \quad 0.9 \quad 1\}$.

The mean-squared error (or residual), namely the error between the output of the optimal model, denoted by $\hat{y}_j^{\text{opt}}(k)$ and given by Eq. (27), and the perturbed outputs $y_j^{e1}(k)$ resulting from the variations of the emulator parameter γ_{j1}. The mean-squared error, denoted $\text{mse}_j(\gamma_{j1})$, is computed as follows:

$$\text{mse}_j(\gamma_{j1}) = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{y}_j^{\text{opt}}(k) - y^{e1}(k)\right)^2 \tag{51}$$

The conventional scheme identifies only the unperturbed nominal model. Let the identified model of the MISO system (10) be $\hat{\boldsymbol{G}}_{j0}(z)$, the estimated output be $\hat{y}_{j0}^{c}(z)$. The mean-squared error, denoted $\text{mse}_j^c(\gamma_{j1})$, becomes

http://dx.doi.org/10.5772/67870

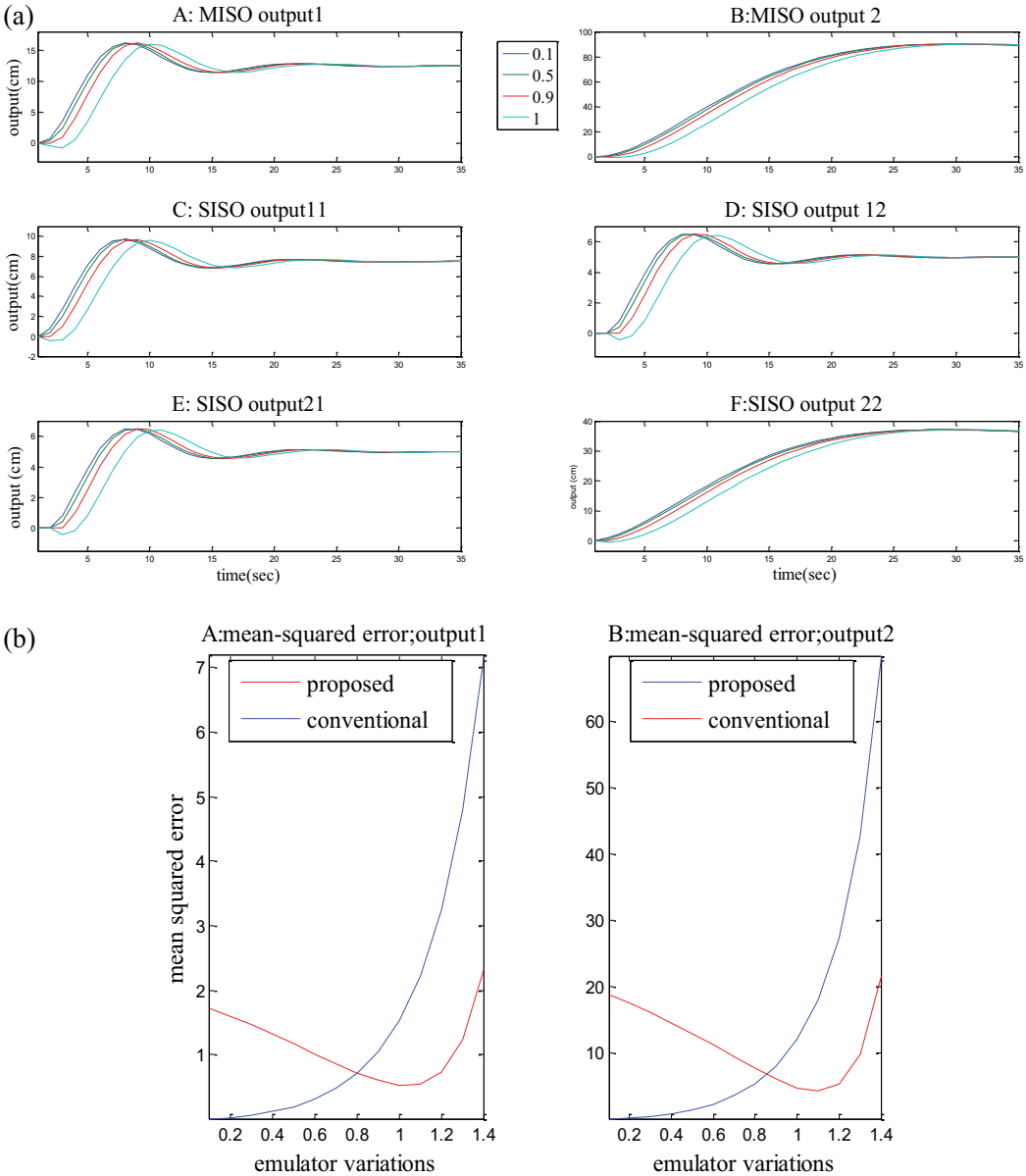


Figure 4. (a) Emulator generated data and (b) performance of the identified model.

$$\text{mse}_j^c(\gamma_{j1}) = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{y}_{j0}(k) - y^{e1}(k)\right)^2 \tag{52}$$

The mean-squared errors $\text{mse}_j(\gamma_{j1})$ and $\text{mse}_j^c(\gamma_{j1})$ are plotted as functions of the emulator parameter perturbations $\Delta\gamma_{j1}$. The mean-squared profiles of both the proposed emulator-based and the conventional identification schemes are shown in subfigures A and B of **Figure 4b**.

The identified state-space model and the Kalman gain are

$$\hat{\boldsymbol{A}}_0 = \begin{bmatrix} 0.9843 & -0.1588 & 0.0213 & 0.0224 \\ 0.1572 & 0.9266 & 0.2317 & -0.2502 \\ -0.0631 & -0.3144 & 0.9090 & -0.3230 \\ 0.0171 & 0.0153 & 0.1659 & 0.7225 \end{bmatrix}, \hat{\boldsymbol{B}}_0 = \begin{bmatrix} 0 & 0 \\ 0.1 & 0.1 \\ -0.3 & -0.3 \\ 2.6 & 2.6 \end{bmatrix} 10^3,$$

$$\hat{\boldsymbol{C}}_0 = \begin{bmatrix} 21 & -564.2 & -245.6 & 49.5 \\ 1679.2 & -336.3 & -211.4 & 44.0 \end{bmatrix}$$

The ranges of the mean-squared errors $\text{mse}_j(\gamma_{j1})$ and $\text{mse}_j^c(\gamma_{j1})$ are given below:

$$\begin{aligned} 1.8390 \leq \text{mse}_1(\gamma_{j1}) \leq\ 2.225 \\ 0.0137 \leq \text{mse}_1^c(\gamma_{j1}) \leq 7.1815 \end{aligned} \tag{53}$$

$$\begin{aligned} 18.2224 \leq \text{mse}_2(\gamma_{j1}) \leq\ 21.7167 \\ 0.0007 \leq \text{mse}_2^c(\gamma_{j1}) \leq 69.8841 \end{aligned} \tag{54}$$

Remarks: The emulator-generated data cover the operating scenarios, including both the normal and abnormal ones, exhibiting variations of the rise time, the settling times, and the overshoots.

The identified optimal model $(\hat{\boldsymbol{A}}_0, \hat{\boldsymbol{B}}_0, \hat{\boldsymbol{C}}_0)$ is different from the nominal system model $(\boldsymbol{A}_0, \boldsymbol{B}_0, \boldsymbol{C}_0)$. Even the block diagonal strictures of $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are not preserved.

It can be deduced from **Figure 4b** on the right, Eqs. (53) and (54), that compared to the conventional scheme, the proposed emulator-based identification is significantly more robust to variations in the operating points, which are simulated by emulator parameter perturbations.

The poles of the MISO transfer functions $\boldsymbol{G}_2(z)$ of $y_2(k)$ and $\boldsymbol{G}_1(z)$ of $y_1(k)$ were, respectively, $0.8500 \pm \text{j}0.3122$ and $0.7500 \pm \text{j}0.3708$. The same emulator was used for inducing phase shift to the MISO models. $\boldsymbol{G}_2(z)$ with poles close to the unit circle was affected more than $\boldsymbol{G}_1(z)$ with poles well inside. In view of the difference in the perturbations induced in the two models, the mean-squared errors mse_2 and mse_2^c are higher than mse_1 and mse_1^c.

7.2. Fault diagnosis

Detection of a fault: Various types of faults include (a) actuator, (b) sensor, and (c) plant, we introduced by varying the columns of $\boldsymbol{B}_0$, the rows of $\boldsymbol{C}_{0,}$ and the diagonal matrices of $\boldsymbol{A}_0$. A fault is detected using appropriate test statistics depending upon the reference input waveform from Eq. (45). Since the reference input $\boldsymbol{r}(k)$ is a constant waveform, the test statistics for the MISO and the SISO system using Eqs. (46) and (47) are

$$t_s(e_j) = \left| \frac{1}{N} \sum_{i=k-N+1}^{k} e_j(i) \right| \ ; \ t_s(e_{ji}) = \left| \frac{1}{N} \sum_{i=k-N+1}^{k} e_{ji}(i) \right| \tag{55}$$

A visual picture of the faulty and the normal subsystems may be deduced from the autocorrelations of the residuals associated with the fault-free, sensor fault, actuator fault, and the plant faults shown in **Figure 5**. Subfigures A, and B, subfigures C and D, subfigures E, and F, and

subfigures G and H show respectively autocorrelations of the residual for the ideal no fault, the sensor fault, the actuator fault, and the plant fault.

Remarks: The maximum value of the autocorrelation of the residual (i.e., its variance) provides an indication of the presence or an absence of the fault. In the case of the sensor fault introduced by perturbing $\boldsymbol{C}_{20}$, it affects only the residual $e_2(k)$. The variance of the autocorrelation $e_2(k)$ is large while that of $e_1(k)$ indicating a fault in $\boldsymbol{C}_2$. However, a fault in either the actuator or the plant, depending upon which elements of $\boldsymbol{B}_0$ or $\boldsymbol{A}_0$ are perturbed, may affect both residuals, and hence would be difficult to isolate.

7.2.1. Fault isolation

If a fault is asserted, and the path where the fault is located, then it is isolated using Bayesian multiple hypotheses testing scheme. The size of the fault is also estimated. The objective of fault isolation is to determine which of the emulator parameter has varied using the residual

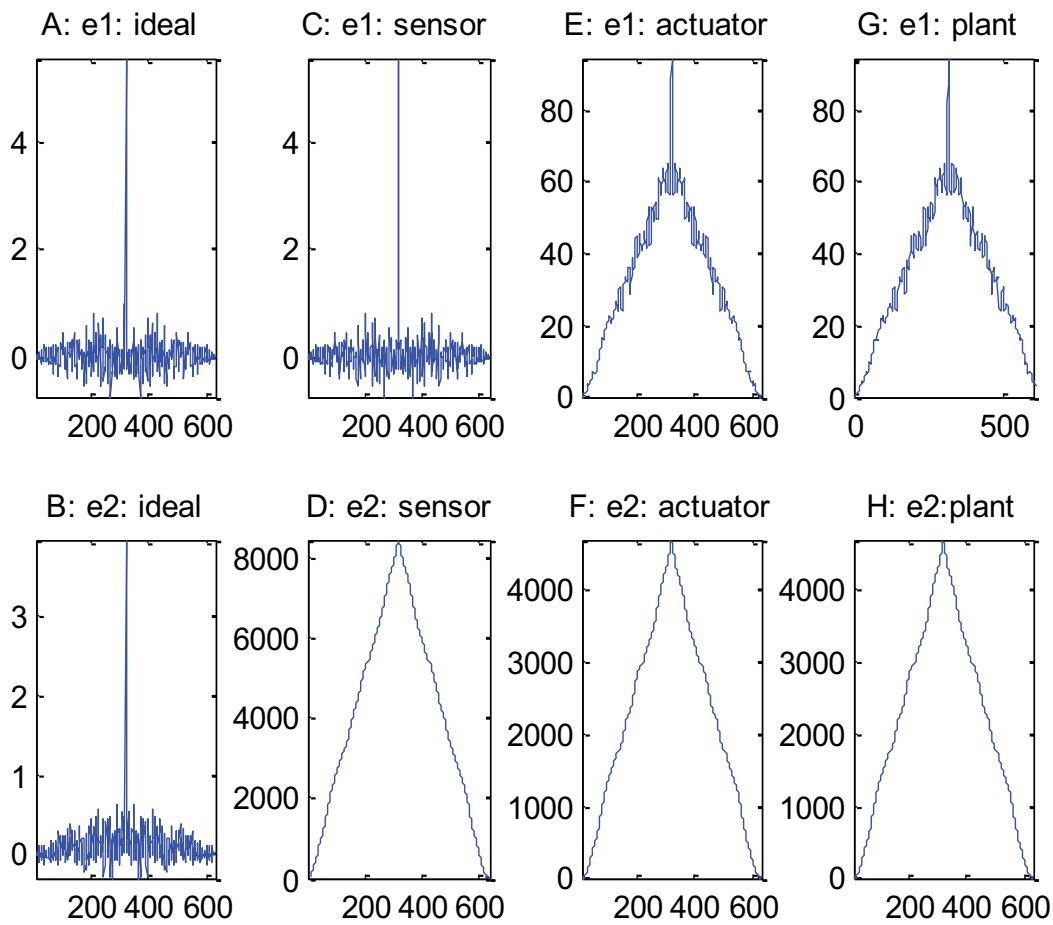


Figure 5. Autocorrelations of the residuals: ideal, sensor fault, actuator fault, and plant faults.

data generated or parameters using the expression for the Kalman filter residual for the model-mismatch case given in Eq. (43). The residual $e_{ji}(k)$ is affine in the unknown emulator parameter variations $\{\Delta\gamma_{ijk}\}$. The emulator parameter variation that is most likely to fit the perturbed residual with additive term $\boldsymbol{\psi}_{jif}^T(z)\Delta\boldsymbol{\theta}_{ji}$ is determined sequentially by first hypothesizing single faults. If the estimates thus obtained do not *fit* the residual, then two simultaneous faults are hypothesized. If again the estimates do not *fit* the residual model, then hypothesize triple faults, and so on until the estimates fit the residual model. The maximum likelihood method, which is efficient and unbiased, is employed herein to estimate the variation $\Delta\boldsymbol{\gamma}$. The maximum likelihood estimates of the emulator parameters are obtained by minimizing the log likelihood function [13].

Let $\boldsymbol{H}^{(1)}$, $\boldsymbol{H}^{(2)}$, and $\boldsymbol{H}^{(3)}$ denote a hypothesis that emulator parameter γ_{ji1}, γ_{ji2}, and γ_{ji12} has varied. The Kalman filter residual for $\boldsymbol{H}^{(1)}$ becomes

$$\boldsymbol{H}^{(1)} : e_{ji}^{(1)}(k) = \boldsymbol{\psi}_{jif}^T(k)\Delta\boldsymbol{\theta}_{ji}^{(1)} + v_{jif}(k) \tag{56}$$

The least-squares estimate $\Delta\hat{\gamma}_{ji1}$ from

$$\Delta\hat{\gamma}_{ji1} = \arg\min_{\{\Delta\gamma_{ji1}\}}\{\|e_{ji}(k) - \boldsymbol{\psi}_{jif}^T(k)\Delta\boldsymbol{\theta}_{ji}\|^2\} \tag{57}$$

If the estimate does not meet the criteria, then hypothesize that γ_{ji2} has varied. The criteria for fitting a hypothesis are given later. The Kalman filter residual for $\boldsymbol{H}^{(2)}$ becomes

$$\boldsymbol{H}^{(2)} : e_{ji}^{(2)}(k) = \boldsymbol{\psi}_{jif}^T(k)\Delta\boldsymbol{\theta}_{ji}^{(2)} + v_{jif}(k) \tag{58}$$

The least-squares estimate $\Delta\hat{\gamma}_{ji2}$ from

$$\Delta\hat{\gamma}_{ji2} = \arg\min_{\{\Delta\gamma_{ji2}\}}\{\|\boldsymbol{\psi}_{jif}^T(k)\Delta\boldsymbol{\theta}_{ji}^{(2)} + v_{jif}(k)\|^2\} \tag{59}$$

If it does not meet the criteria, then hypothesize that γ_{ji12} has varied. The Kalman filter residual for $\boldsymbol{H}^{(3)}$ becomes

$$\boldsymbol{H}^{(3)} \;:\; e_{ji}^{(3)}(k) = \boldsymbol{\psi}_{ijf}^T(k)\Delta\boldsymbol{\theta}_{ji}^{(3)} + v_{jif}(k) \tag{60}$$

The least-squares estimate $\Delta\hat{\gamma}_{ji12}$ from

$$\Delta\hat{\gamma}_{ji12} = \arg\min_{\{\Delta\gamma_{ji12}\}}\left\{\| e_{ji}^{(3)}(k) - \boldsymbol{\psi}_{ijf}^T(k)\Delta\boldsymbol{\theta}_{ji}^{(3)}\|^2\right\} \tag{61}$$

where $\Delta\boldsymbol{\theta}_{ji}^{(1)}$, $\Delta\boldsymbol{\theta}_{ji}^{2}$, and $\Delta\boldsymbol{\theta}_{ji}^{3}$ are deviations in the feature vector when γ_{ji1}, γ_{ji2}, and γ_{ji12} are assumed to have varied.

http://dx.doi.org/10.5772/67870

7.2.2. Criteria for asserting the hypothesis

The most likely hypotheses is determined by verifying which of the emulator parameter or parameters have varied by comparing the deviation with some threshold value

$$Assert \quad \boldsymbol{H}^{(1)} \ \ if \ \ \Delta\hat{\gamma}_{ji1} \geq \eta_1 \tag{62}$$

$$\text{Assert} \quad \boldsymbol{H}^{(2)} \ \ if \ \ \Delta\hat{\gamma}_{ji2} \geq \eta_2 \tag{63}$$

$$\text{Assert} \quad \boldsymbol{H}^{(3)} \ \ if \ \ \Delta\hat{\gamma}_{ji12} \geq \eta_3 \tag{64}$$

where η_1, η_2, and η_3 are threshold values. The subsystem associated with the subsystem is asserted to be faulty if the criterion is met.

8. Evaluation on physical process control system

A laboratory-scale two-tank physical system is formed of a controller, a DC motor, a pump, two tanks connected by a pipe, a flow rate sensor, and a liquid level sensor. The system is interfaced to a PC with the National Instruments LABVIEW for data acquisition and implementing the controller and the soft sensor [14]. The actuator, namely the pump driven by the DC motor, sends the fluid to the first tank to maintain a specified fluid level in the second tank. An evaluation of the proposed scheme for fault diagnosis was performed on a benchmark laboratory-scale process control system using the National Instruments LABVIEW as shown below in **Figure 6**. The sampling period is $T_s = 0.05$.

Emulator-generated height and flow rate profiles under various types of faults are shown in under the caption Height/Flow rate Profiles for PI controller with Consumer in **Fig. 7**. **Figures 7a–c** show the height and flow rate profiles when subjected to (a) leakage fault, (b) actuator fault, and (c) sensor faults, respectively. The height profile is shown on the top and the flow rate profile is shown at the bottom of **Figure 7**. The faults are induced by varying the appropriate

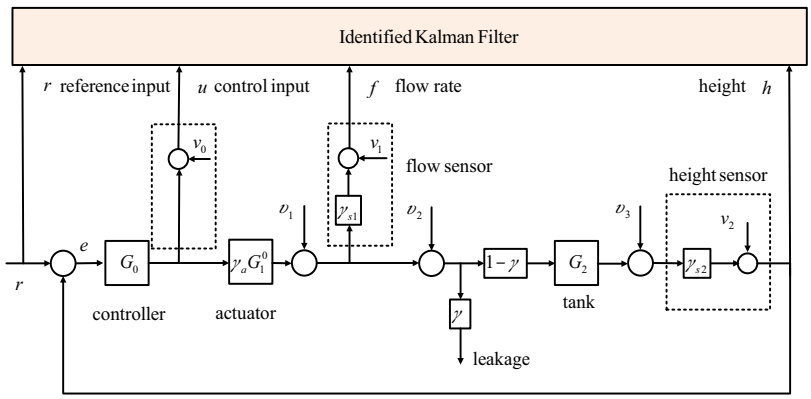


Figure 6. Process control system: controller, actuator, and tank.

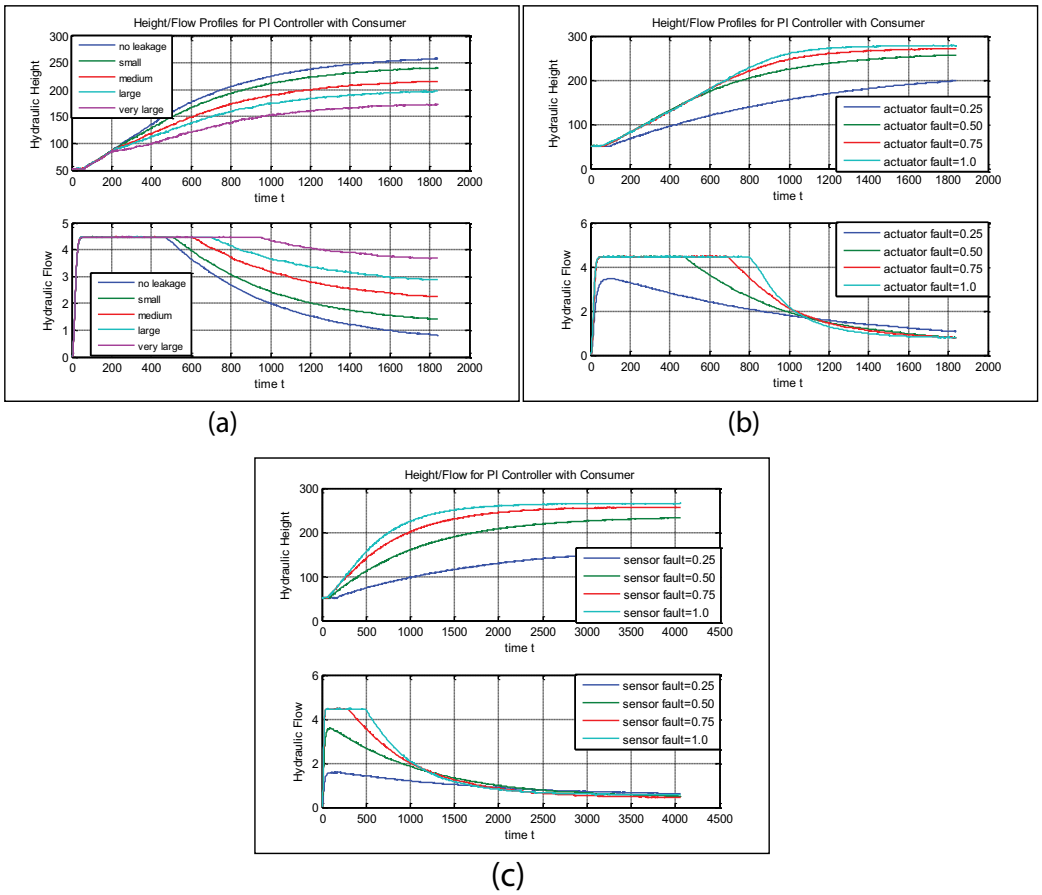


Figure 7. Emulator-generated data: height and flow.

emulator parameters to 0.25, 0.5, and 0.75 time the nominal values, in order to represent "small," "medium," and "large" faults. However, by virtue of its control design objective, the closed-loop PI controller will hide any fault that may occur in the system and hence will make it difficult to detect it. In addition, the physical system exhibits a highly nonlinear behavior. The flow rate saturates at 4.5 ml/s. The dead-band effect in the actuator exhibits itself as a delay in the output response: when a step reference input is applied, the height output responds after some delay, as a minimum force is required to drive the actuator. These nonlinearities affect the steady-state value of the height: even though there is an integral action in the closed-loop control system, the steady-state error is non-zero for a constant reference input.

The system is modeled as a single-input, multi-output system where r is the reference input, and the outputs are the control input u, the flow rate f and the height h. Faults were induced in the height sensor, the flow sensor, the actuator, and also as a leakage. The proposed fault diagnosis successfully detected and isolated all the faults compared to SISO scheme [14], where all the faults were detected and isolated using the reference input and the height output.

http://dx.doi.org/10.5772/67870

9. Conclusions

Fault detection and isolation of a class of linear multiple-input and multiple-output system based on the Kalman residual and the emulators were presented. The key properties of the Kalman filter, namely the residual, is a zero-mean white noise process if and only if there is no model mismatch, drive the prediction error identification of the nominal system model, and the Kalman filter. In view of the closed-loop configuration, the noise and the disturbance are attenuated at the estimated output. The Kalman filter is the best minimum variance estimator in the class of all linear estimators.

To handle fault isolation, the powerful and effective concept of emulators was introduced. Similar in spirit to the training of the artificial neural network, a number of emulator parameter-perturbed experiments were performed to capture the perturbation model of the subsystems to help with fault isolation. The influence vectors of the emulator parameters, which are indirectly the associated subsystems, were estimated. The influence vectors captured the emulator perturbation model and hence that of the subsystem.

The residual of the Kalman filter was shown to have an additive fault indicating term when there is a model mismatch due to emulator perturbations. The model-mismatch term is affine in the emulator parameter variations. Using the expression for the fault indicating term, the fault was isolated using the influence vectors and its size was estimated. The residual, being affine in the emulator parameter variation, easily lends itself to the widely used and successful composite Bayes hypothesis-testing scheme for fault isolation.

The future work generated from this work includes its extension to a class of nonlinear multiple-input and multiple-output systems, and the development of a computationally efficient identification of the Kalman filter directly from the input data even for unstable systems. Although a gold standard for system identification, the prediction error method involves a nonlinear optimization problem and hence can suffer from the existence of local minima. Unlike the least-squares approach, it does not offer a closed-form solution to the parameter estimation problem. Instead, it relies on a recursive solution that may be time-consuming (slow convergence rate), computationally complex, and which may also suffer from initialization problems.

Acknowledgements

The first author acknowledges support of the Department of Electrical and Computer Engineering, the University of New Brunswick, and the National Science and Engineering Research Council (NSERC) of Canada. The author is grateful to Professor C.P. Diduch of the University of New Brunswick, Mr. Jiong Tang of MDS and Mr. H.M. Khalid of KFUPM for their help and suggestions. The second author acknowledges the support of KFUPM, Saudi Arabia, and the help of Mr. H. M. Khalid of KFUPM.

Author details

Rajamani Doraiswami[1] and Lahouari Cheded[2]*

*Address all correspondence to: cheded@kfupm.edu.sa

1 Department of Electrical and Computer Engineering, University of New Brunswick, Fredericton, NB, Canada

2 Systems Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

References

[1] Doraiswami, Rajamani, and Lahouari Cheded. "Kalman Filter for Fault Detection: An Internal Model Approach." *IET Control Theory and Applications* 6, no. 5 (2012): 1–11.

[2] Ljung, Lennart. *System Identification: Theory for the User.* New Jersey: Prentice-Hall, 1999.

[3] Doraiswami, Rajamani, and Lahouari Cheded. "A Unified Approach to Detection and Isolation of Parametric Faults Using a Kalman Filter Residuals." *Journal of Franklin Institute* 350, no. 5 (2013): 938–965.

[4] Doraiswami, Rajamani, Chris Diduch, Maryhelen Stevenson. *Identification of Physical Systems: Applications to Condition Monitoring, Fault Diagnosis, Soft Sensor and Controller Design.* John Wiley and Sons, ISBN 9781119990123, 2014.

[5] Doraiswami, Rajamani, and Lahouari Cheded. "Linear Parameter Varying Modelling and Identification for Condition-based Monitoring of Systems." *Journal of Franklin Institute* 352, no. 4 (2015): 1766–1790.

[6] Haykin, Simon. *Neural Networks: A Comprehensive Foundation.* New Jersey: Prentice Hall, 1999.

[7] Ding, SX. *Model-Based Fault Diagnosis Techniques: Design Schemes.* Springer-Verlag, London, ISBN 978-3-540-76304-8, 2008.

[8] Gertler, Janos, F. *Fault Detection and Diagnosis in Engineering Systems.* Marcel-Dekker Inc, New York, ISBN 0-8247-9427-3, 1998.

[9] Isermann, Rolf. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance.* Springer-Verlag, 2006.

[10] Patton, Ronald, J, Paul, M Frank, and Robert, N Clark. *Issues in Fault Diagnosis for Dynamic Systems.* Springer-Verlag, 2000.

http://dx.doi.org/10.5772/67870

[11] Pertew, AM, HJ Marquez, and Q Zhao. "H-infinity Observer Design with Applications in Fault Diagnosis." Seville, Spain, 2006. 3803–3809.

[12] Silvio, Simani, Cesare Fantuzzi, and Ronald, J Patton. *Model-based Diagnosis using Identification Techniques.* Advances in Industrial Control, Springer-Verlag New York, Secaucus, NJ USA,1852336854, 2003.

[13] Doraiswami, R, C Diduch, and J Tang. "A New Diagnostic Model for Identifying Parametric Faults." *IEEE Transactions on Control System Technology* 18, no. 3 (2010): 533–544.

[14] Doraiswami, Rajamani, L Cheded, and M.H Khalid. *Sequential integration approach to fault diagnosis with applications: model-free and model-based approaches.* VDM Verlag Dr. Muller Aktiengesselschaft and Co, 2010.

Chapter 2

Adaptive Signal Decomposition Methods for Vibration Signals of Rotating Machinery

Wei Guo and Ming J. Zuo

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67530

Abstract

Vibration-based condition monitoring and fault diagnosis are becoming more common in the industry to increase machine availability and reliability. Considerable research efforts have recently been directed towards the development of adaptive signal processing methods for fault diagnosis. Two adaptive signal decomposition methods, i.e. the empirical mode decomposition (EMD) and the local mean decomposition (LMD), are widely used. This chapter is intended to summarize the recent developments mostly based on the authors' works. It aims to provide a valuable reference for readers on the processing and analysis of vibration signals collected from rotating machinery.

Keywords: signal processing, empirical mode decomposition, local mean decomposition, fault diagnosis, rotating machinery

1. Introduction

Signal processing methods with adaptive basis functions are more effective in revealing the overlapping components in vibration signals. They are able to adaptively disassemble nonlinear and non-stationary signals into some simpler signal components. The empirical mode decomposition (EMD) [1] method and the local mean decomposition (LMD) [2] method have been recognized to be such effective adaptive signal processing methods.

Since the introductions of EMD in year 1998 [1] and LMD in year 2005 [2], many improvements and applications have been reported. In this chapter, we summarize the recent developments mostly based on the authors' works. We hope that it is a valuable reference for readers on the processing and analysis of vibration signals collected from rotating machinery. This chapter

is organized as follows. Section 2 briefly introduces the fundamentals of EMD and LMD. Section 3 summarizes key results on the improvements of EMD and LMD. Section 4 outlines future work and remaining challenges.

2. Fundamentals of EMD and LMD

The EMD method [1] decomposes a nonlinear and non-stationary signal as a sum of some intrinsic mode functions (IMFs). Resembling the popular wavelet transform, EMD can also display the spread of signal energy on available frequencies locally in time [3]. Their key difference is that the EMD method is direct and adaptive, so that some potential and valuable information can be obtained from the data without the influence from *a priori* basis. Hence, it is widely applied in diverse areas of signal processing, especially in the field of mechanical vibration, such as health monitoring and diagnosis, analysis and identification of weak vibration signals, mainly in rotating machinery with critical elements, bearings and gears [3].

Another adaptive signal processing method, the LMD method, was originally used as a time-frequency analysis tool of the electroencephalogram signals [2]. It is an iterative approach to decompose a signal into some product functions (PFs) [2], each of which is an amplitude-modulated and frequency-modulated signal (AM-FM signal) from which the instantaneous amplitude (IA) and the instantaneous frequency (IF) can be derived [4]. Compared with the corresponding IMF in EMD, as shown in **Figure 1**, the calculated IF and IA are not involved in the EMD and require adopting of the Hilbert transform (HT).

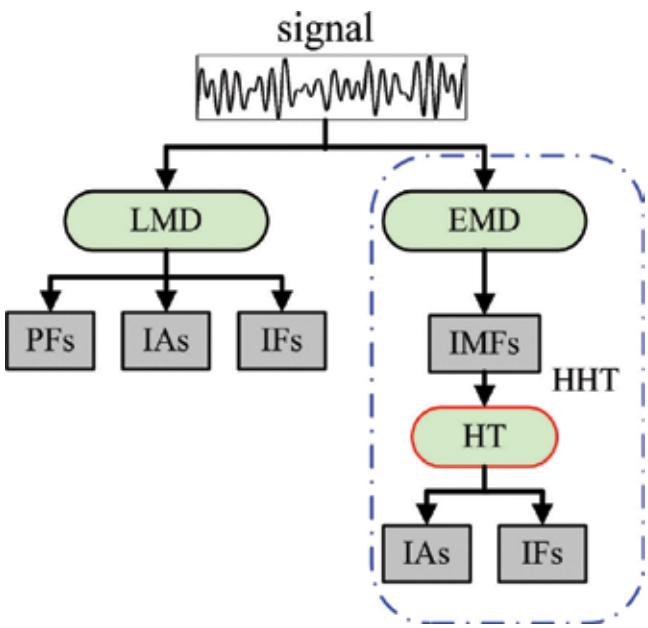


Figure 1. Main differences between EMD and LMD [4].

Although the details of decomposition and the resulted signals are quite different, these two methods share some common advantages, for example, the adaptive property. They also share some common challenges, which will be addressed in Section 3. Ref. [4] provides a comparative study, and Ref. [5] reviews applications of EMD in the field of fault diagnosis.

No matter which of two methods is used, a multi-component signal, $x(t)$, can be adaptively decomposed into k mono-components, $x_p(t)$ (p = 1, 2, …, k) (IMFs for EMD or PFs for LMD) and a residue, u_k, and can be reconstructed by summing them together, i.e.

$$x(t) = \sum_{p=1}^{k} x_p(t) + u_k(t). \quad (1)$$

3. Reported improvements in EMD and LMD

The EMD and the LMD methods are proven to be quite versatile in a broad range of applications for adaptively extracting signals of interest from noisy data. This section discusses their main and common challenges, including end effects, mode mixing, feature signal selection and strong noise reduction. After analysing each issue, the corresponding improvement is also shown. Other open issues, such as stopping criterion and envelope function, will be briefly discussed in Section 4.

3.1. End effects

End effects have plagued data analysis from the beginning of any known method [6]. The end effects were first mentioned in the spine fitting of the EMD. This section briefly reviews related improvements and then introduces an adaptive method to eliminate the end effects for the vibration signals collected from rotating machinery.

3.1.1. Improvements for eliminating end effects

Two ways have been proposed to eliminate end effects. One timid way is to use a sliding window [7], as is done routinely in Fourier analysis [6]. The sliding window is successfully applied to Fourier analysis using various windows and continuous wavelet analyses. However, appropriate and reliable windows are often analysis method related but not related to the data themselves. It inevitably leads to sacrifice some precious data near the ends [8]. Furthermore, it would be a hindrance for data processing when the data are short.

The other elimination way is extension or prediction of data beyond their existing range, which is still the best basic solution. Huang et al. [1] first proposed to add characteristic waves to treat the effects, in which the extra points are determined by the average of n-waves in the immediate neighbourhood of the ends. Motivated by this idea, some extension methods tried to extend the temporal waveforms forward and backward by using all available information in data, including feature-based extension, mirror or anti-mirror extension, intelligent prediction, pattern comparison, etc. [9].

It is proven that prediction methods can provide good performance on the extension of data. It is not needed to predict the whole time series, but to predict the value and location of the extrema adjacent to the ends. However, as pointed out by Huang and Shen [6], the data extension or prediction is a risky procedure even for linear and stationary processes. For nonlinear and non-stationary processes, the problems, such as predictable conditions, method and accuracy, are still open at present. Meanwhile, intelligent methods have their own shortcomings, including minima and over-fitting in artificial neural network (ANN) and sensitiveness to parameter selection in both support vector regression and ANN.

No matter which method is developed, their main idea is that newly added points have minimal interior perturbations and extend the signal implicitly or explicitly beyond the existing range. Furthermore, the extending data can well repeat the form or feature of the original signal. The reliability of such extension will sharply decrease as its distance away from the known data set increases, and thus it is necessary to be careful in extending a signal only by adding the extrapolation data to it [10]. Otherwise, the error of such operation would propagate from the end to the interior of the data and even cause severe deterioration of the whole signal [9].

For most of the vibration signals generated by rotating machinery, their non-linear and non-stationary properties are definite, which is quite challenging for data extension. Although the mirror image extension is easier to be put into practice, the real case that the data are mostly from non-stationary stochastic systems must be faced. Fortunately, the vibration signal has an advantage to assist the extension: it is cyclo-stationary [11]. Meanwhile, the extension based on characteristics of the signal waveform seems to be more appropriate to describe such complexity of problems [10]. In the following section, an adaptive waveform extension method [9] is introduced to extend vibration signals and avoid error accumulation.

3.1.2. Adaptive data extension-based spectral coherence

To facilitate applications to condition monitoring and fault diagnosis, the designed extension method should have good extension performance as well as easy operation to implement. An adaptive extension method [9] was designed for vibration signals, mainly including three steps: waveform segmentation, spectral coherence comparison and waveform extension. Its main idea is to automatically search inside waveforms having mostly similar frequency spectrum to ends, and then use their successive segments for signal extension. In this method, a critical point is how to measure the waveform similarity. Although there are some similarity measures, such as correlation coefficient, cross-correlation, waveform similarity, originally used in the field of data fusion, pattern recognition and speech recognition, most of them are susceptible to noise and not suitable for processing vibration signals since their acquisition and transmission often suffer from noise. Therefore, an index measuring the spectral coherence [12] is introduced here. The procedure is described as follows [9]:

Step 1. *Waveform segmentation*. Identify zero crossings of the analyzed signal and then separate the signal into N segments, $c_i(t)$ ($i = 1, \cdots, N$).

Step 2. *Segment repetition and fast Fourier transform (FFT)*. Repeat each segment to form a long waveform and then conduct fast Fourier transform (FFT).

http://dx.doi.org/10.5772/67530

Step 3. *Spectral coherence comparison*. Use Eq. (2) to calculate the revised spectral coherence (RSC) values of the first segment $c_1(t)$ and other segments, and then find the segment, denoted as $c_{\text{back}}(t)$, having the largest value of the RSC. Similarly, search the segment $c_{\text{for}}(t)$ similar to the last segment $c_N(t)$.

$$\gamma_{i,j} = \frac{\sum_F C_i(F) \times C_j(F)}{\sqrt{\left(\sum_F C_i(F) \times C_i(F)\right)\left(\sum_F C_j(F) \times C_j(F)\right)}}, \tag{2}$$

where $C_i(F)$ and $C_j(F)$ are frequency spectra of the signal components $c_i(t)$ and $c_j(t)$, respectively.

Step 4. *Waveform extension*. Use the pervious segment of $c_{\text{back}}(t)$ for extending backward and the next segment of $c_{\text{for}}(t)$ for extending forward.

Based on this, the extended signal can be decomposed by the EMD or LMD method, and extended samples are at last truncated before further analysis. Using the hidden periodicity, a cyclo-stationary signal, for example, a vibration signal, can be easily extended beyond its original range, and its temporal continuity in time domain and spectral coherence in frequency domain can be properly maintained.

3.1.3. Experiment and analysis

A vibration signal collected from an industrial traction motor [13] is shown in **Figure 2**. The specification of the experiment setup is given in **Table 1**. This signal is cyclo-stationary with around three cycles, and its waveform is thus divided into six segments, which are marked in **Figure 2**. **Figure 3** shows frequency spectra of these segments.

To estimate the influence caused by the end effects, a measure of energy change [14] before and after decomposition is defined as

$$\theta = \frac{1}{R_x}\left|\sqrt{\sum_{p=1}^{k} R_p^2 + R_u^2} - R_x\right|, \tag{3}$$

where R_x, R_p and R_u are root-mean-square (RMS) values of the original signal $x(t)$, the pth product function $\text{PF}_p(t)$, and the residue signal $u_k(t)$, respectively. The value of the measure is $\theta \geq 0$. The closer the measure is to zero, the smaller the error between the original signal and decomposition results is; that is to say, the influence caused by end effects is smaller.

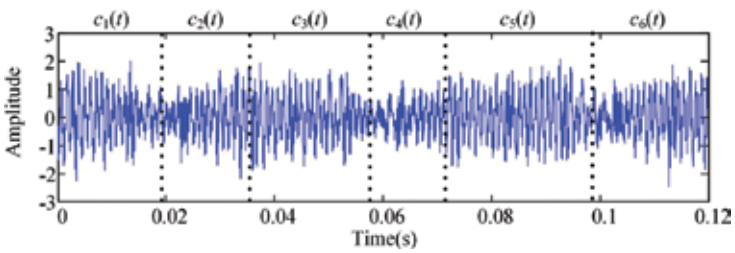


Figure 2. A raw vibration signal collected from a traction motor.

Section	Tested object	Fault type	f_s (kHz)	f_r (Hz)	f_d (Hz)
3.1.3 3.2.2 3.4.3.2	Bearing	An outer race defect on bearing, and motor eccentric	32.8	25	BPFO = 114
3.2.1	Bearing	An outer race defect on bearing	80	23.3	BPFO = 135
3.3	Bearing	An inner race defect on bearing	80	23.3	BPFI = 192
3.4.3.1	Bearing	An outer race defect on bearing	51.2	60	BPFO = 183

Note: f_s—sampling frequency; f_r—rotating frequency of the motor; f_d—characteristic defect frequency; BPFO—the ball pass frequency of the outer race; BPFI—the ball pass frequency of the inner race.

Table 1. Specifications of the experiment setup.

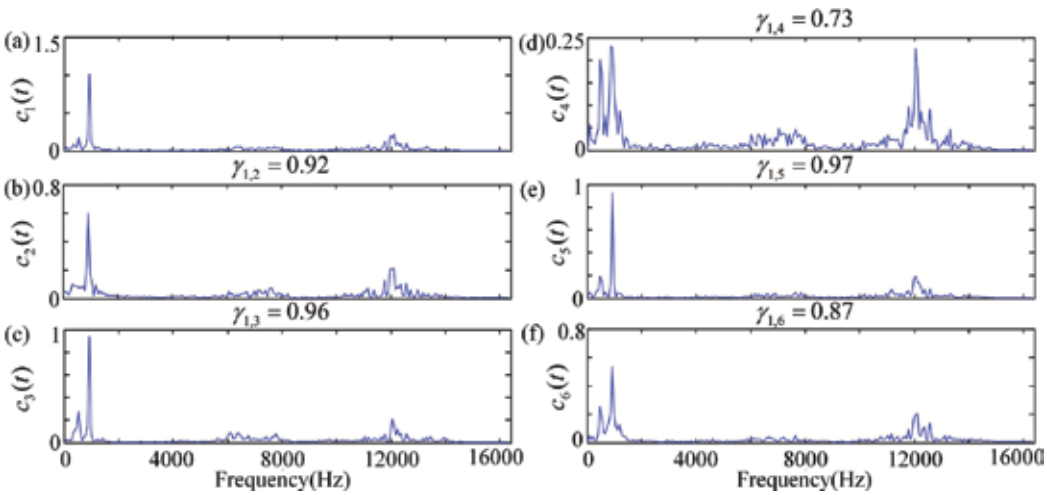


Figure 3. Frequency spectra of six segments $c_1(t)$-$c_6(t)$ in (a)–(f) in **Figure 2**, and spectral coherence values between $c_1(t)$ and the other segments.

The revised spectral coherence (RSC) values $\gamma_{1,j}$ (j = 2, …, 6), i.e. RSC values between the segment $c_1(t)$ and one of other segments, are shown above each sub-figure in **Figure 3**. It can be seen that the segment $c_5(t)$ has the largest RSC value of 0.97 and its previous segment $c_4(t)$ is then used for backward extension of $c_1(t)$. In a similar way, $c_2(t)$ has the largest RSC value with the last segment $c_6(t)$, and thus the next one of $c_2(t)$, i.e. $c_3(t)$ is used for forward extension. The extended vibration signal is shown in **Figure 4**, where extended waveforms are shown in red. Its RSC value with the original signal is 0.94, and the measure θ is 0.005. If no extension, the measure θ is 0.106.

http://dx.doi.org/10.5772/67530

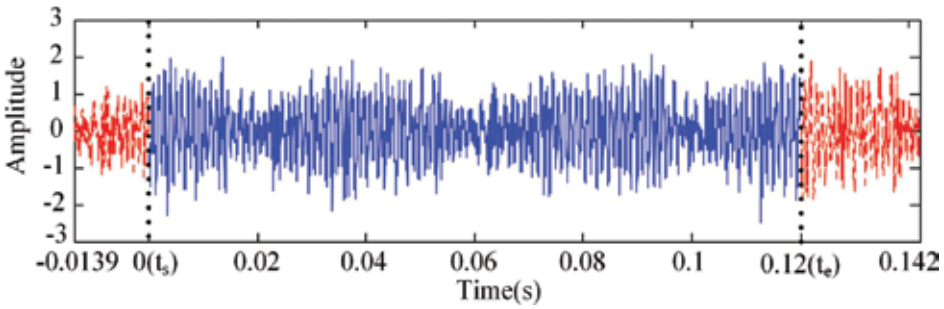


Figure 4. The extended vibration signal obtained by the adaptive waveform extension method [9].

After applying the LMD method to the extended signal and truncating extended parts, five PFs and a residue are obtained, the first three of which have larger correlation coefficient values with the original signal and thus are selected for further analysis. Their waveforms and envelope spectra are shown in **Figure 5**. In **Figure 5(d)**, the identified characteristic frequency (104 Hz) and its harmonics (around 2 × and 3 × BPFO) can be easily observed. The error between the theoretical value (114 Hz) and the identified one (104 Hz) is mainly caused by inaccurate shaft speed after long use and limited samples (only 0.12 second). In **Figure 5(e)** and **(f)**, higher impulses are identified at the frequency of 25 Hz, corresponding to the motor rotating frequency. It indicates that PF1 is the signal generated by the inspected bearing with an outer race defect, and PF2-3 is generated by the motor, which turned out to be caused by the eccentric problem after inspection. More cases on bearings and gears can be found in Ref. [9].

3.2. Mode mixing

Another open problem for EMD and LMD is the mode mixing. It is originally defined as a single IMF either consisting of signals of widely disparate scales, or a signal of a similar scale residing in different IMF components, which causes serious aliasing in the time-frequency

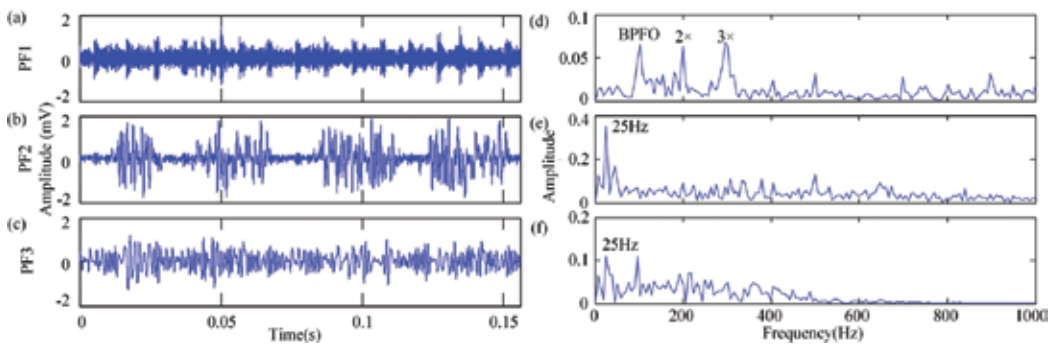


Figure 5. Decomposition results PF1-PF3 in (a)–(c) by applying the LMD to the signal in **Figure 4**, and their corresponding envelope spectra in the range of 0–1 kHz in (d)–(f).

distribution and makes the meaning of IMF unclear [8]. This section focuses on the solution to the problem of mode mixing.

3.2.1. Separation of disparate components

According to the above definition, there are two possibilities: either completely different components existing in one IMF, or one component appearing in more than one IMF. To remove the former case, Wu and Huang [8] presented a noise-assisted signal processing method, called ensemble EMD (EEMD). In this method, white noise with a pre-setting amplitude is introduced to perturb the analyzed signal and enables the EMD method to visit all possible solutions in the finite neighbourhood of the true final IMF [8]; and the ensemble means of decomposition results help to remove the remaining noise in the results. For the EEMD method, two parameters, the noise amplitude and the ensemble number, are critical, the former of which has more influence on its performance [15]. In order to process signals adaptively, it is ideal to automatically find appropriate parameters for the analyzed signal. A parameter optimization method [13] is designed for the EEMD. In this method, an index termed relative root-mean-square error (RMSE) is first used to evaluate the performance of the EEMD method when fixing a small ensemble number and setting various noise amplitudes, and then the signal-to-noise ratio (SNR) is introduced to evaluate the remaining noise in the results when gradually increasing the ensemble number.

For a signal, $x_o(k)$, it is assumed that it consists of main component(s), background noise and some components having small correlation coefficients with the chief one, which has the largest correlation coefficient with the signal $x_o(k)$ is marked as $c_{\max}(k)$. The desired decomposition is to completely separate the component $c_{\max}(k)$ from others, and the relative RMSE is thus used to evaluate the separation performance when setting various noise amplitudes. Its formulation is

$$\text{Relative RMSE} = \sqrt{\frac{\sum_{k=1}^{S}(x_o(k)-c_{\max}(k))^2}{\sum_{k=1}^{S}(x_o(k)-\bar{x}_o)}}, \tag{4}$$

where $\bar{x}_o$ is the mean of the signal $x_o(k)$, and S is the number of samples in this signal. The value of this index is in the range of 0–1. The smaller this index is, the closer the component $c_{\max}(k)$ to the original signal. It means that the extracted IMF contains not only the main component of interest but also other components, and thus the objective is not achieved. However, there exists a value for the noise amplitude that maximises the index. At this point, the error between $x_o(k)$ and $c_{\max}(k)$ is from noise and other components, that is to say, the extracted IMF and the other in the original signal share no common component, and the main component of interest is extracted from the original signal. The corresponding value is the optimal noise amplitude. Its procedure is briefly described as follows [13]:

Step 1. Set a small value of the initial ensemble number, for example, $N_E = 10$, and choose a relatively large value as the initial noise level, $L_N = l_0$. The noise amplitude A is to multiply the noise level by the standard deviation of the signal.

Step 2. Perform the signal decomposition using the EEMD method and calculate the relative RMSE of the chief component $c_{\max}(k)$.

Step 3. Decrease the noise level and repeat Step 2 until the change in the relative RMSE is negligible or small enough.

Step 4. Identify the optimal noise level corresponding to the maximal relative RMSE.

Once the optimal noise level is numerically determined, the ensemble number can be determined by comparing the SNR values when gradually increasing the ensemble number from its pre-setting value.

To demonstrate this method, a vibration signal was collected from a small motor [13] and is shown in **Figure 7(a)**. In the experiment, a fault was set on the outer race of the tested bearing. The specification of the experiment is shown in **Table 1**. Initial parameters are set as: a larger value for the noise level, $L_N = 2$, and the ensemble number $N_E = 10$. During the execution of the above program, the noise level is gradually decreased. When $2 \le L_N \le 0.1$, the noise level is decreased in the step of 0.1; when $0.1 < L_N \le 0.01$, its decreasing step is 0.01; when $0.01 < L_N \le 0.001$, its decreasing step is 0.001.

After applying the EEMD method with the above optimization method to decompose the vibration signal, the relative RMSEs for various noise levels are shown in **Figure 6**. As shown in this figure, the maximal relative RMSE is arrived at the noise level of 0.4, corresponding to the optimal one, and accordingly, the extracted IMF (IMF1) is shown in **Figure 7(b)**. Comparing with the original signal, most of noise and redundant components are separated from IMF1, and its kurtosis value is 26.07.

To compare with this, extracted IMFs when setting any three non-optimal noise levels are also shown in **Figure 7**. **Figure 7(c)** and **(d)** shows the results when setting the noise levels of 2 and 1, respectively. **Figure 7(e)** shows the result when setting a quite small noise level of 0.009. Their kurtosis values of these IMFs are 11.68, 7.26 and 7.65, respectively. It demonstrates that better decomposition results are obtained after setting the optimal noise level.

Having determined the optimal noise level, appropriate ensemble number is then determined. The variation in the SNR is shown in **Figure 8**. As the figure shows, when the ensemble number is less than 80, increasing the ensemble number gently accelerates the increase in the SNR value. When the ensemble number is larger than 120, the SNR value fluctuates smoothly.

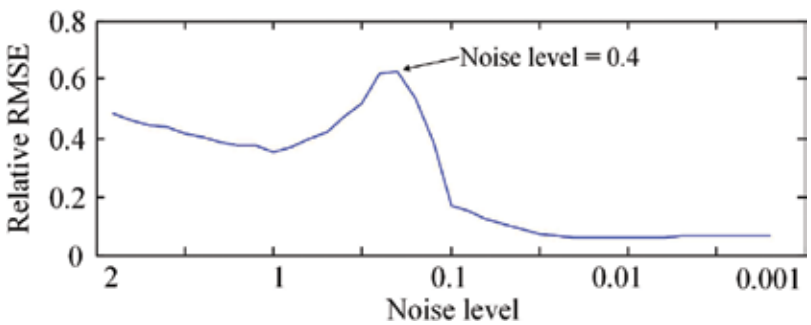


Figure 6. Relative RMSEs when adding white noise with various noise levels to the vibration signal in **Figure 7(a)**, and the optimal noise level for this signal is 0.4.

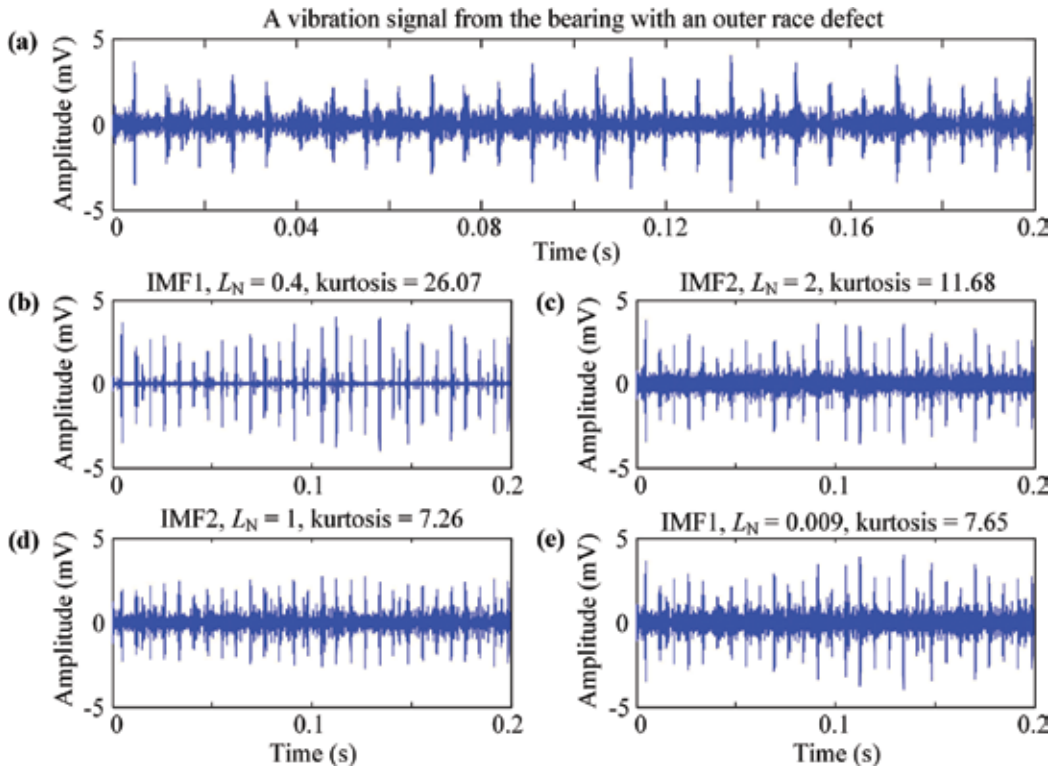


Figure 7. A vibration signal from the bearing with an outer race defect and the corresponding selected IMFs when setting the optimal and three non-optimal noise levels. (a) The signal to be analyzed; (b) IMF1 when setting LN = 0.4; (c) IMF2 when setting LN = 2; (d) IMF2 when setting LN= 1; and (e) IMF1 when setting LN = 0.009.

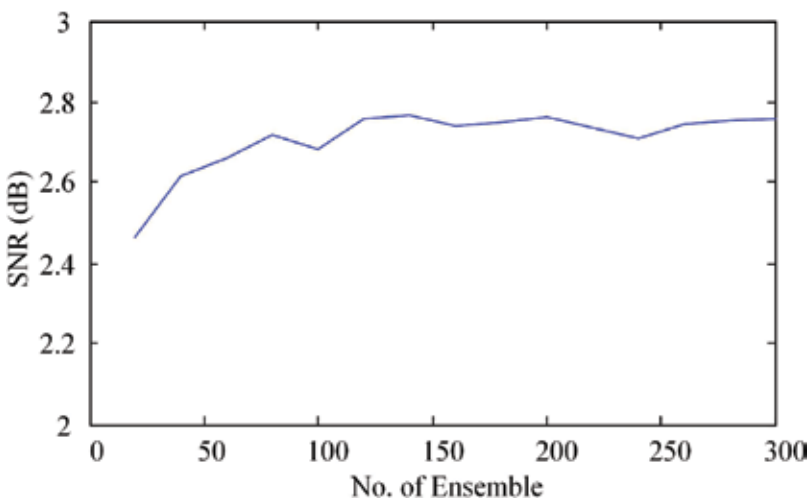


Figure 8. SNR values for various ensemble numbers when setting the optimal white noise (L_N = 0.4) to the vibration signal in **Figure 7(a)**.

Further increasing its value contributes to minor increasing of the SNR, but definitely rising computation cost. Therefore, using this optimization method, parameters of the EEMD can be automatically determined according to the signal itself, instead of empirical setting or the trial and error. More cases on bearings can be found in Ref. [13].

3.2.2. Mixing of similar components

Although the EEMD method can successfully separate signal components with different scales, another mode mixing still exists in the decomposition results, that is to say, one component may spread in more than one IMF. This also belongs to the mode mixing and results in energy dispersion and some redundant components without physical significance. It may be caused by repeated sifting process and severe stopping criterion. A simple and convenient solution is to combine the components from the same source. Therefore, the index of spectral coherence in Eq. (2) is used to evaluate the spectral similarity of two successive components and then combine the components with similar spectra into a natural IMF [12].

Using the index of spectral coherence, the similarity criterion of two successive IMFs obtained by the EEMD method is described as:

(1) If $\gamma_{j,j+1} \to 1$, it means that the IMFs, c_j and c_{j+1}, have a relationship of similarity in frequency domain, that is to say, they have spectral coherence over the whole frequency range. Thus, these two IMFs should come from the same source and thus are combined to one natural IMF (NIMF).

(2) If $\gamma_{j,j+1} \to 0$, they have low, even no spectral coherence and thus are two natural IMFs.

(3) If $\gamma_{j,j+1}$ is around 0.5, the spectral coherence of two IMFs cannot be determined. Such signal components are also viewed as two natural IMFs and would not be combined together.

The signal in **Figure 2** is used to demonstrate the process of similarity analysis and combination. After applying the EEMD method with the noise level 0.2 and the ensemble number of 30 to the signal, 12 IMFs are obtained, the first four of which have larger correlation coefficient values with the original signal and are shown in **Figure 9**. As shown in the figure, the frequency spectrum of IMF1 is a high-frequency dominated signal and centred at the frequency of 12 kHz, and it indicates that IMF1 corresponds to the signal generated by the faulty bearing in the traction motor; as for IMF3 and IMF4, they share the common frequency of 920 Hz generated by the faulty motor. Furthermore, the revised spectral coherences of all IMFs are calculated and the results are shown in **Table 2**. According to this table, there are three local minimal points, i.e. $\gamma_{2,3}$, $\gamma_{5,6}$ and $\gamma_{7,8}$. The RSC values of IMF3-IMF4 and IMF4-IMF5 are larger than 0.5 and it shows their similarity on frequency domain, and thus these three components are combined to one natural IMF. Between the second and the third local minimal values, IMF6 and IMF7 show the spectral similarity. Similarly, the remaining components, IMF8-IMF12, also show their spectral similarity, and thus are merged into another natural IMF. The RSC value of IMF1 and IMF2 is not close to 1 or 0 and these two IMFs are thus two natural IMFs. Final results are shown in **Figure 10**. The last two components are practically residues. Based on local minima of RSC, a fusion rule [12] was designed to automatically combine

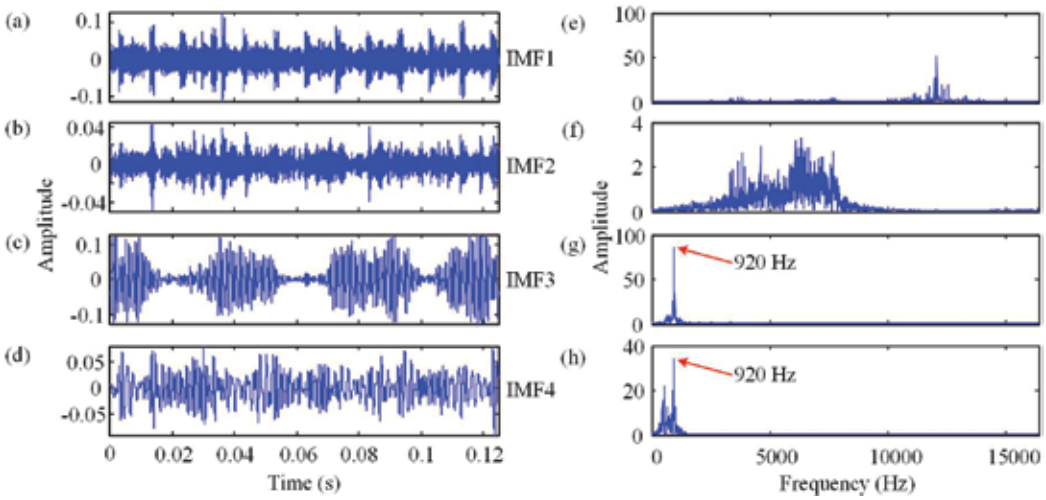


Figure 9. IMF1-IMF4 obtained by applying the EEMD (N_E = 30, L_N = 0.2) to the signal in **Figure 2** and frequency spectra of these IMFs (a)-(d) The temporal waveforms of IMF1-IMF4, and (e)-(h) corresponding frequency spectra of IMF1-IMF4 in (a)-(d).

RSC	$\gamma_{1,2}$	$\gamma_{2,3}$	$\gamma_{3,4}$	$\gamma_{4,5}$	$\gamma_{5,6}$	$\gamma_{6,7}$	$\gamma_{7,8}$	$\gamma_{8,9}$	$\gamma_{9,10}$	$\gamma_{10,11}$	$\gamma_{11,12}$
Value	0.279	**0.054**	0.768	0.542	**0.375**	0.568	**0.398**	0.593	0.619	0.993	0.958

Table 2. RSC values of decomposition of the signal in **Figure 2** by using EEMD (N_E = 30, L_N = 0.2).

components from the same source and remove the mode mixing in the original EMD method. Other applications on bearings can be found in Ref. [12].

3.3. Strong noise reduction

In real rotating machinery, a raw vibration signal generally consists of strong noise and two or more sources. Some vibrations, such as improper installation and surfacing of the installed sensors, random impacts from friction and contact forces and external disturbances [16], are also so strong that the signal of interest is completely overwhelmed. Therefore, the recovery of the feature signal from noise, while preserving its features is a challenging problem. This section introduces a hybrid signal processing method [17] for noisy vibration signals.

3.3.1. Problem analysis

Although the EEMD method improves the scale separation ability of EMD method, both methods are based on extrema to discriminate signals generated by various sources. When the signal of interest is completely overwhelmed by strong noise, there may be a lack of necessary extrema for the EEMD method to separate the real signal from noise. An experimental signal collected from a bearing with an inner race defects is used to illustrate this problem [17]. The specification of the experiments is shown in **Table 1**. To simulate strong noise in real

http://dx.doi.org/10.5772/67530

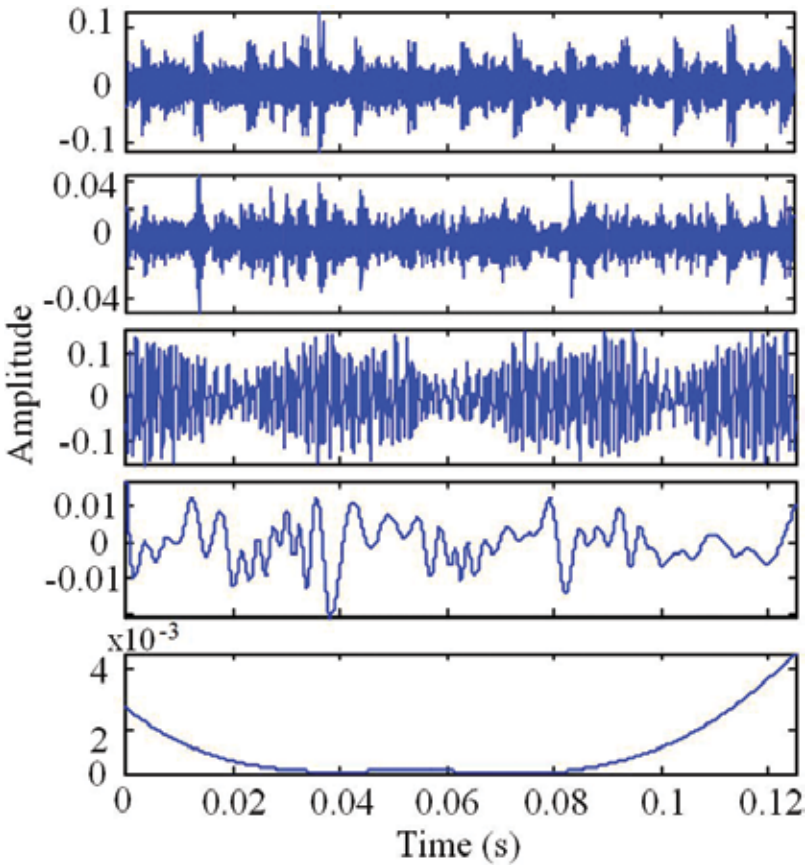


Figure 10. Five natural IMFs after applying similarity criterion to the results obtained by the EEMD.

cases, Gaussian white noise was added to the experimental signal, and the generated noisy signal is shown in **Figure 11**. As shown in the figure, the impulses caused by faulty bearing are completely masked by strong noise. After applying the EEMD method to this signal, 13 IMFs are obtained and the first four having larger correlation coefficient values with the original signal are shown in **Figure 12**, in which impulses are seldom observed and still buried in noise. It is because that the decomposition method lacks necessary extrema generated by the tested faulty bearing.

As for a signal with a relatively low signal-to-noise ratio, it is necessary to design an adaptive filter to extract the weak feature signal of interest from a noisy signal to facilitate further signal decomposition. A possible solution is to use the spectral kurtosis, which is proven to be a powerful tool to identify the existing of bearing faults buried in noise. Its value is large in frequency bands where the impulsive bearing fault signal is dominant, and is effectively zero where the spectrum is dominated by stationary components [16]. Based on this, an SK-based filter [18] was used to pre-process the signal in **Figure 11** and remove part of noise before decomposition. It is a kind of band-pass filter whose parameters, centre frequency and bandwidth,

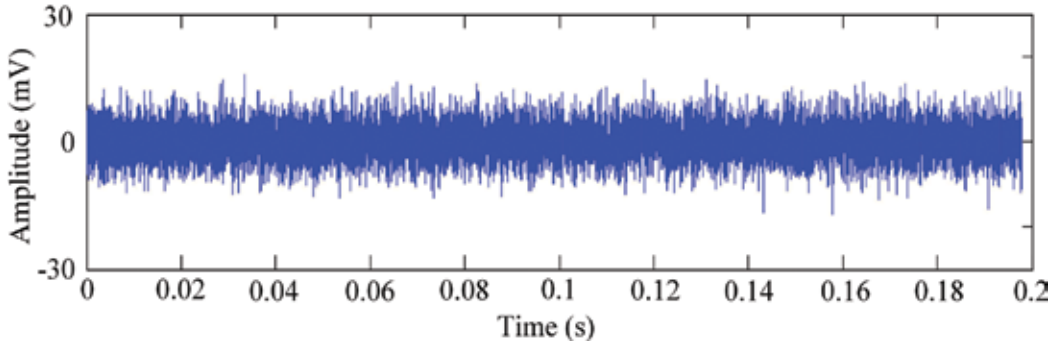


Figure 11. A vibration signal from a bearing with an inner race defect with the added white noise.

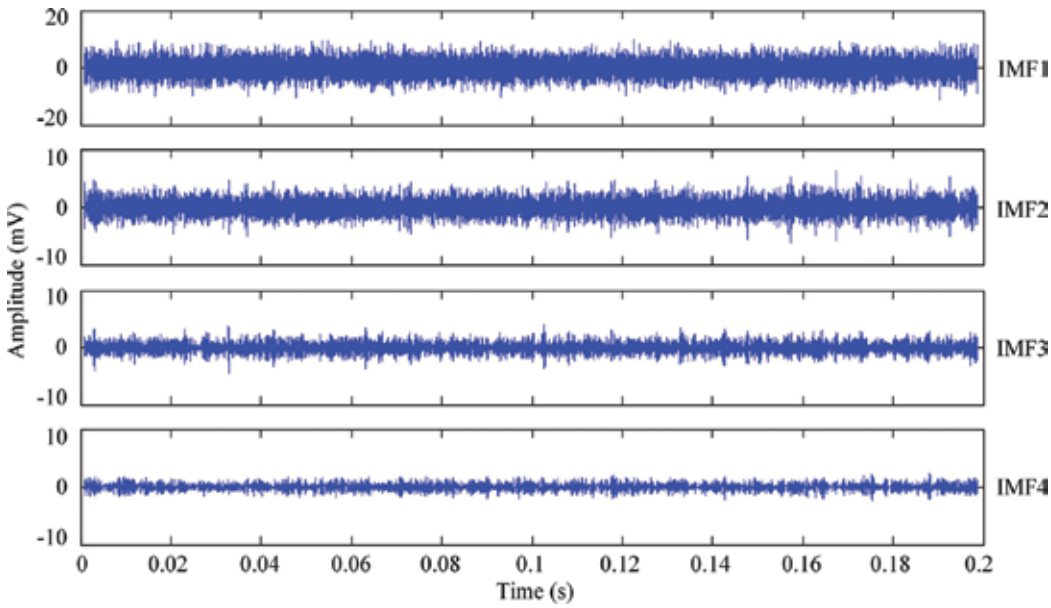


Figure 12. The first four IMFs obtained by applying the EEMD method to the signal in **Figure 11**.

are optimized by using the kurtogram, a map formed by the STFT-based (short-time Fourier transform based) SK. The filtered result is shown in **Figure 13**. Although the filtered signal still contains some noise, its impulses are a little clearer than those in the original signal, and its kurtosis value is also increased from 3.07 to 3.97. Consequently, a hybrid method is used to reinforce the performance of noise reduction.

3.3.2. A hybrid method for strong noise reduction

By comparing individual performances of the foregoing two methods, a hybrid signal processing method that combines the EEMD and the SK-based filter [17] is introduced. First, an optimal band-pass filter based on SK is employed to remove part noise so that local extrema

http://dx.doi.org/10.5772/67530

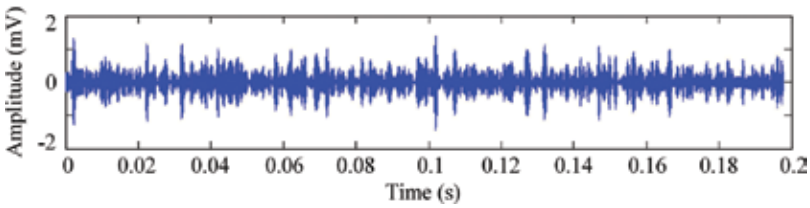


Figure 13. The filtered signal using the SK-based filter, and its kurtosis value is 3.97.

of the signal would not be completely concealed by noise. Then, the EEMD method with parameter optimization is applied to further decompose the filtered signal. As a result, the final signal can be separated from strong noise, which allows good detection of the defects but at the same time minimizes the distortion of the impulses. The main procedure is as follows:

Step 1. *Pre-processing*. Filter the raw signal using an optimal band-pass filter based on SK and obtain the filtered signal.

Step 2. *Signal decomposition*. Use the EEMD method to decompose the filtered signal into some IMFs.

Step 3. *Selection of feature signal*. Calculate the correlation coefficients between the obtained IMFs and the filtered signal, and select the IMF having the largest values of correlation coefficient (CC) as the resultant signal for further analysis.

3.3.3. Experiment and comparison

In this sub-section, the filtered signal in **Figure 13** is decomposed into 13 IMFs, the first three of which have CC values of 0.88, 0.76 and 0.18 with the filtered signal, and the rest of which have CC values close to zero. To save space, only the first four IMFs are shown in **Figure 14**. According to the calculation results of CC, IMF1 has a larger correlation coefficient (0.88) than the other signal components and contains the main component in the filtered signal, and it is thus viewed as the bearing signal recovered from the noisy experimental signal. This result is also verified by the identified BPFI and its multiples as shown in **Figure 15**. There is also an error between the theoretical and identified values of BPFI, which is caused by the same reason mentioned in Section 3.1.3.

Compared with the filtered signal in **Figure 13**, the extracted bearing signal (IMF1 in **Figure 14**) is much cleaner than the original signal, and the remaining noise in the filtered signal is almost completely separated and resides in IMF2. The kurtosis values of the raw signal, the filtered signal and IMF1 are 3.07, 3.97 and 11.29, respectively, as observably increasing. It indicates that this hybrid method successfully reveals temporal impulses from a noisy signal while preserving its important feature for accurate fault diagnosis. **Figure 16** also shows the filtered signal by applying the normal wavelet threshold denoising to the same noisy signal, and its impulses are not as clear as those in **Figure 14**. More cases on faulty machine components, such as an outer race and a rolling ball, are given in reference [17].

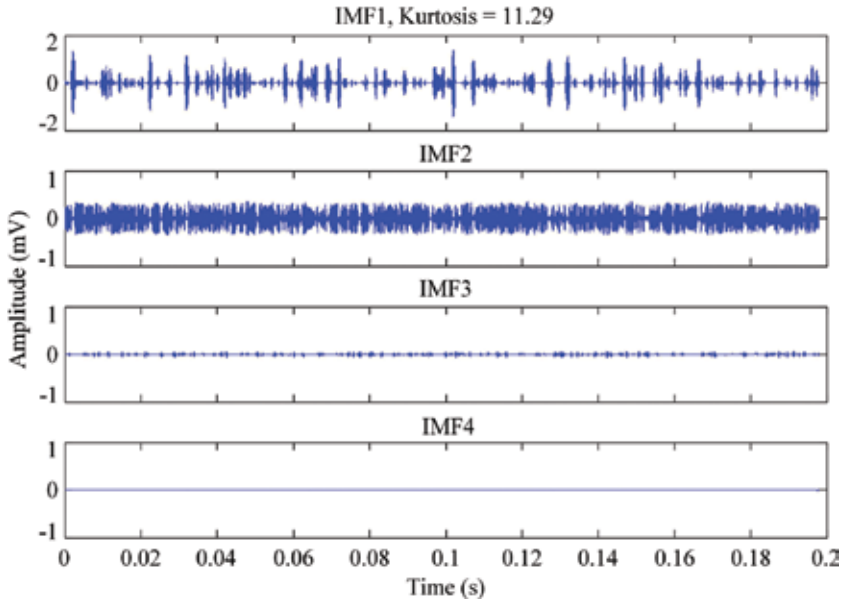


Figure 14. The first four IMFs obtained by applying the hybrid method to the signal in **Figure 11**.

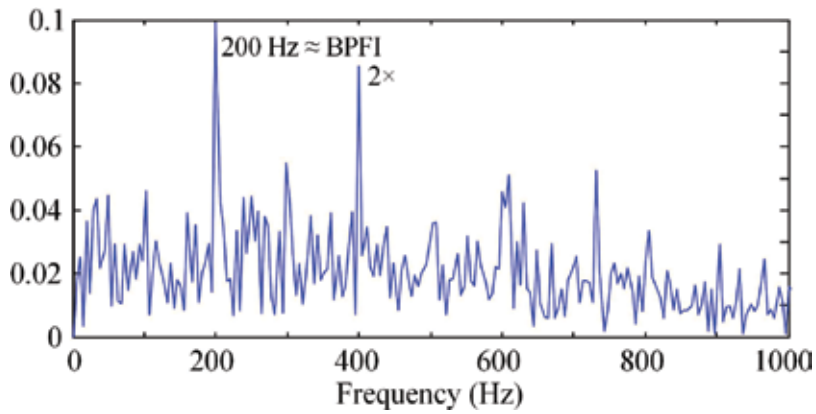


Figure 15. The envelope spectrum of IMF1 shown in **Figure 14** in the range of 0–1 kHz.

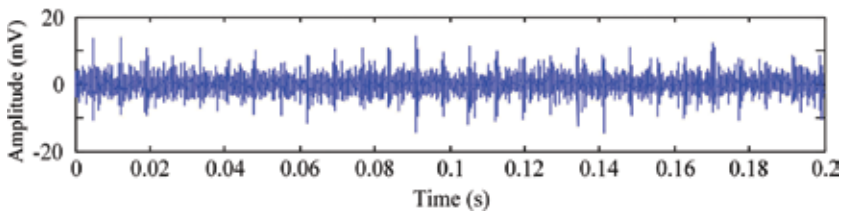


Figure 16. The filtered signal obtained by applying the normal wavelet threshold denoising to the same signal in **Figure 11**.

http://dx.doi.org/10.5772/67530

3.4. Feature signal component selection

After using the EMD or LMD method, many signal components are disassembled from the original signal. How to effectively select feature signals from many components is critical for further signal processing and analysis. This section primarily discusses the selection method of feature signal components.

3.4.1. Selection based on cluster analysis

For the feature signal selection, a popular solution is to calculate statistical indicators of the signal, for example, correlation coefficient (CC). Dybała and Zimroz [19] used this indicator to divide IMFs into three classes: noise-only IMFs corresponding to low indices and low CC values, signal-only IMFs and trend-only IMFs corresponding to high indices and low CC values. However, it is possible that an impact signal caused by a damaged bearing is wrongly categorized into the class of noise [19]. Similar results can be found when only using single measurement. A more sophisticated diagnostic method is needed to avoid the misdiagnosis. Referring to the idea of the cluster analysis, an adaptive selection method based on multiple statistic indicators is designed for selecting the feature signal of interest from many signal components [20].

In the anomaly detection, a branch of cluster analysis, a detector is designed to detect any object that deviates from the known state (usually the healthy state) [21]. Referring to this, the decomposed signal components are classified into two groups: feature signals and unrelated signals. The former is used for further analysis, and the latter is viewed as useless signals.

The key of this selection is how to evaluate useful content in the analyzed signal. If the feature signal is wrongly classified into the useless part, the state of the monitored object may be misjudged. If an unrelated signal is wrongly marked as the feature signal, the conclusion based on the analyses of feature signals may be conflicting. To correctly classify them, some statistic indicators commonly used in the anomaly detection and feature extraction are introduced here. They are indicated by many literatures to be good at representing hidden features of the analyzed signal. Therefore, these indicators are jointly used to determine the classes of decomposed components, not to determine the fault types of the tested object. In addition, the strategy of using multiple indicators is very common in pattern recognition to combine various experts with the aim of compensating the weakness of each single expert [22]. This combination can be viewed as a kind of ensemble learning and can improve the classification accuracy in machine learning. What is interesting is that the idea of combining individuals' opinions in order to reach a final decision is humans' second nature before making any crucial decision [23].

As for a large number of indicators, the distance evaluation technique (DET) [24] is introduced to quickly organize the classification result of each indicator (or expert). For more than one expert, their conclusions may not always coincide, and thus the principle of minority obeying majority [22, 23] is introduced to solve their conflicts. The detailed selection is described in the following sub-section.

3.4.2. Adaptive feature signal selection

The process of the adaptive feature signal selection can be divided into two stages: classification of each expert and decision of all experts. Its procedure is described as follows:

Step 1. Calculate some statistics indicators in time and frequency domains for all decomposed signal components. The indicators include peak-to-peak (P-P), mean, absolute mean, max, root mean square (RMS), standard deviation (SD), skewness, kurtosis, crest factor (CF), shape factor (SF), impulse factor (IF), energy and correlation coefficient (CC).

Step 2. Normalize and sort in a descending order for each indicator.

step 3. Classify using the DET. For each indicator, the DET makes the distance within a class shorter and the distance between classes longer, and then the components are classified into two groups.

Step 4. Vote by all 'experts'. For each signal component, summary how many 'experts' (indicators) classify it into the same class.

Step 5. Draw a conclusion. Following the principle of minority obeying majority, the classification results of signal components can be finally determined.

Furthermore, the indicators that win in the voting are viewed as sensitive ones. After comparing the values of any sensitive indicator between the current state and the healthy one, signals in the class having obvious changes can be determined as feature signals.

3.4.3. Experiments and analyses

3.4.3.1. Case 1: a vibration signal collected from a bearing with single defect

One of experimental signals was collected from a small motor that involves a bearing with an outer race defect [17]. The specification of experiments is shown in **Table 1**. After applying the LMD method to this signal, five PFs were obtained, and then indicator values of these five signal components are calculated by 13 indicators in time domain and another 13 indicators in frequency domain. **Figure 17** shows the indicator values after normalization. As shown in this figure, for the first indicator P-P (peak-to-peak), using the DET, PF3 and other PFs are classified into two groups; while, for the indicator of Mean, PF5 and other PFs belong to different groups. The classification results for all indicators in time and frequency domains are shown in the columns 'Case 1' of **Table 3**. Based on the majority principle, PF1 and PF2 are finally classified into one class, and the rest of PFs are classified into the other class. Comparing the energy value with that of a healthy bearing, PF1 and PF2 are determined as feature signals of interest. To verify this conclusion, envelope spectra of PF1-PF3 are shown in **Figure 18**. The characteristic defect frequency f_d and its multiples are only identified in the spectra of PF1 and PF2, which demonstrates the right selection of feature signals.

http://dx.doi.org/10.5772/67530

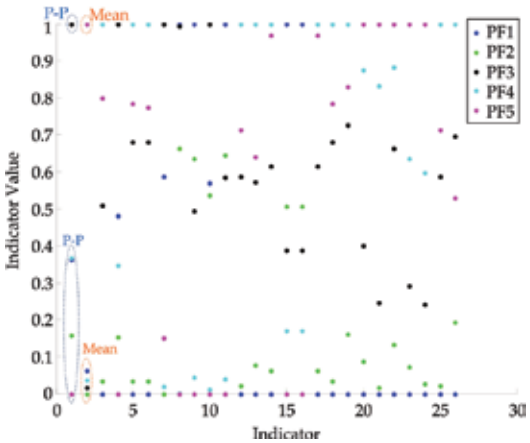


Figure 17. Indicator values of the signal in Case 1 after normalization.

Indicator	Case 1 (T)		Case 1 (F)		Case 2 (T)		Case 2 (F)	
	Class 1	**Class 2**	**Class 1**	**Class 2**	**Class 1**	**Class 2**	**Class 1**	**Class 2**
P-P	3	12 *45*	**345**	**12**	**123**	**4567**	**123**	**4567**
Mean	5	12 *34*	*2* 345	1	2	*13* 4567	12	*3* 4567
Absolute mean	**345**	**12**	*2* 345	1	**123**	**4567**	12	*3* 4567
Max	3	12 *45*	**345**	**12**	**123**	**4567**	**123**	**4567**
RMS	**345**	**12**	**345**	**12**	**123**	**4567**	**123**	**4567**
SD	**345**	**12**	**345**	**12**	**123**	**4567**	**123**	**4567**
Skewness	*1* 3	2 *45*	45	12 *3*	123 *467*	5	123 *45*	67
Kurtosis	45	12 *3*	45	12 *3*	23	*1* 4567	123 *45*	67
CF	45	12 *3*	**345**	**12**	12 *5*	*3* 467	123 *67*	45
SF	45	12 *3*	5	12 *34*	2 *45*	*1 3*67	1234 *57*	6
IF	45	12 *3*	5	12 *34*	12 *5*	*3* 467	1234 *57*	6
Energy	**345**	**12**	**345**	**12**	**123**	**4567**	**123**	**4567**
CC	**345**	**12**	**345**	**12**	**123**	**4567**	**123**	**4567**

Notes:

(1) Title line: T—time domain and F—frequency domain.

(2) Numerical values listed above represent the indices of signal components PFs. For example, in the row of Max, the value '3' in the column of Case 1 (T) means that PF3 is classified into Class 1, and the values '1245' in the column of Case 1 (F) means that PF1, PF2, PF4 and PF5 are classified in to Class 2.

(3) Numerical values in bold indicate that the corresponding PFs are correctly classified after the final voting. And numerical values in Italic indicate that the corresponding PFs are wrongly classified. For example, if using the index of Max in Case 1(T), PF4 and PF5 should be classified into Class 1, not Class 2, and thus the numerical values '45' in the column of Class 2 are italicized.

Table 3. Classification results of two experimental vibration signals.

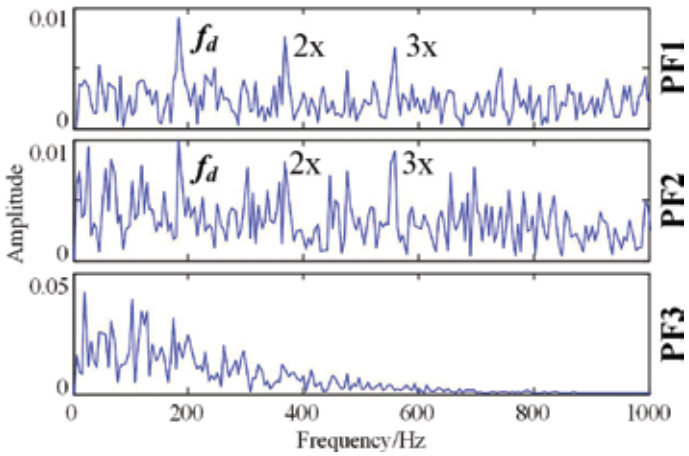


Figure 18. Envelope spectra in the range of 0-1 kHz of part components in Case 1.

3.4.3.2. Case 2: a vibration signal collected from a machine with two defects

Another vibration signal was collected from a traction motor, which involves two faulty machine components, i.e. a faulty motor and a bearing with an outer race defect [13]. Its specification is also shown in **Table 1**. This signal was decomposed into seven PFs by using the LMD. Its classification results are shown in the columns 'Case 2' of **Table 3**. As a result, PF1, PF2 and PF3 are classified into one class, and others belong to another class. **Figure 19**

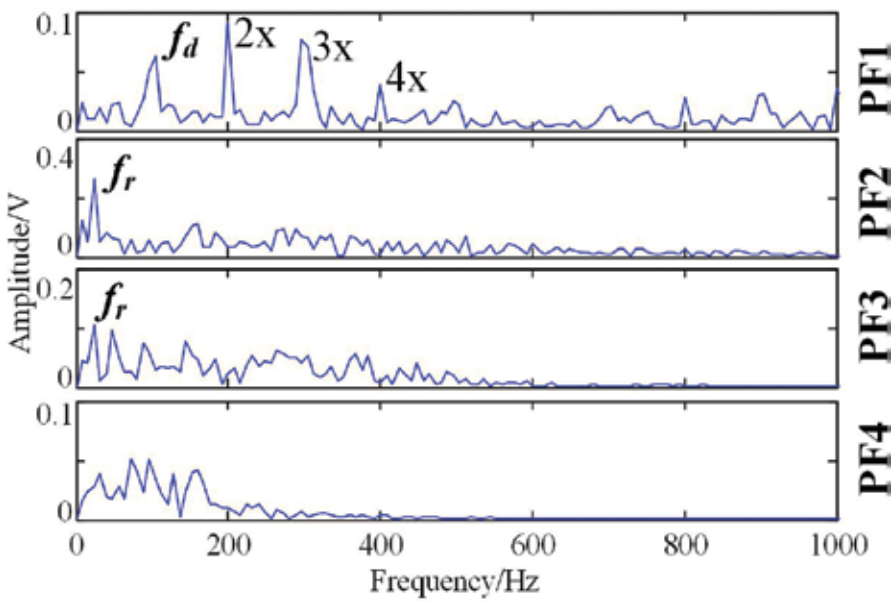


Figure 19. Envelope spectra in the range of 0–1 kHz of part components in Case 2.

shows envelope spectra in the range of 1 kHz of the first four PFs. As shown in this figure, the characteristic defect frequency f_d and its multiples of the faulty bearing are identified in PF1, and the rotating frequency f_r is identified in PF2 and PF3. No specific characteristic frequency can be identified in PF4. These results also match with real condition of the tested machine.

3.4.3.3. Discussion

Above results also indicate that statistic indicators have varying degrees on sensitivity to abnormal states. Some of them are sensitive and closely related to any faults, but others are not sensitive or stable. For above experiments, sensitive indicators include absolute mean, SD, RMS, energy, correlation coefficient in time domain, and max, peak-to-peak, SD, RMS, energy, correlation coefficient in frequency domain. The commonly used indicator, kurtosis in time and frequency domains does not show its sensitivity to feature signals. Although the indicator Energy is one of sensitive features, its values of five PFs in Case 1 are 0.048, 0.053, 0.192, 0.293 and 0.223, the latter three of which corresponding to useless components are much larger. Therefore, single measure is not suitable for fault detection. Further work on assessment of feature signals is necessary for online monitoring and diagnosis.

4. Future work

Although EMD and LMD methods are quite simple in principle, they also depend on a number of user-controlled tuning parameters and still lack an exact theoretical foundation. Feldman has given some theoretical analyses of the EMD method in Refs. [3, 25]. However, the following issues remain to be further addressed.

4.1. Stopping criterion

No matter which method, EMD or LMD, you use, the adaptive signal decomposition is a 'sifting' process, and you need to choose a criterion to stop it at the right time, which is critical for signal processing. The more times sifting is taken, the closer to zero the average will be [26], that is to say, by sifting as many times as possible, it is more likely to eliminate the riding waves and make the wave profiles symmetric. However, too many repetitions would result in the obliteration of the amplitude variation and the loss of physical meanings. Therefore, it is not an easy task to define an appropriate criterion that makes the definition of IMFs satisfied while retaining enough physical sense of amplitude and frequency modulations.

Standard stopping criterion is very rigorous and difficult to implement in practice. The most commonly used criterion is three-threshold criterion [27], and the recommended setting for the three thresholds is applicable for most of the cases. Many modifications on this criterion are also reported, and their wide verifications are not yet finished. Since most of stopping criteria are the summations over the global domain, an undesired feature is that

the decomposition is sensitive to the local perturbation and to the addition of new data [8]. Therefore, an open problem is to eliminate extra sifting processes cause by local changes.

4.2. Connection between local extrema

In the sifting process of the EMD method, a spline interpolation function is needed to connect the identified local extrema. Commonly used spline functions include linear spline, quadratic spline, cubic spline and cubic Hermite spline (third-order polynomial). Generally, the higher-order spline function can provide better fitting performance for the original signal, whereas, they require additional subjectively determined parameters and take considerable time for computation. The selection of spline function should satisfy the least interferences and maximum smoothness.

Similarly, smoothed connecting between successive extrema is also required to form a smoothly varying continuous function in the LMD method, and the parameter selection of the moving averaging is still explored. Although modifications based on single connection method or a hybrid method are sporadically reported, an appropriate criterion on the selection of connection methods receives little attention and remains an open problem.

Considering that the EMD and the LMD are data-driven analysis methods, they are essentially algorithmic in nature and, hence, suffer from the drawback that there is no well-established analytical formulation on the basis of theoretical analysis and performance evaluation [28]. Accordingly, relevant modifications mainly come from case-by-case comparisons conducted empirically. In spite of this, as adaptive signal processing methods, the EMD and the LMD methods are proven to be useful and adaptive signal processing tools for vibration-based fault diagnosis and detection.

Acknowledgements

The work described in this chapter is fully supported by the National Natural Science Foundation of China (Nos. 51405063, 51375078 and 51537010), and China Postdoctoral Science Foundation Funded Project (No. 2016M590874).

Author details

Wei Guo[1]* and Ming J. Zuo[1,2]

*Address all correspondence to: guo.w@uestc.edu.cn

1 School of Mechatronics Engineering, University of Electronic Science and Technology of China, Chengdu, China

2 Department of Mechanical Engineering, University of Alberta, Edmonton, Canada

References

[1] Huang N.E., Shen Z., Long R., Wu C., Shih H., Zheng Q., Yen C., Tung C., Liu H. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineer Sciences. 1998; 454: 903–995.

[2] Smith J.S. The local mean decomposition and its application to EEG perception data. Journal of the Royal Society Interface. 2005; **2**(5): 443–454.

[3] Feldman M. Analytical basics of the EMD: two harmonics decomposition. Mechanical Systems and Signal Processing. 2009; **23**(7): 2059–2071.

[4] Wang Y.X., He Z.J., Zi Y.Y. A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis. Journal of Vibration and Acoustics—Transactions of the ASME. 2010; **132**(2): 021010 (10 pages).

[5] Lei Y., Lin J., He Z.J., Zuo M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing. 2013; **35**(1–2): 108–126.

[6] Huang N.E., Shen S.S.P. Hilbert-Huang Transform and Its Applications. World Scientific, Singapore; Hackensack, NJ; London; 2005.

[7] Pai P.F., Palazotto A.N. Detection and identification of nonlinearities by amplitude and frequency modulation analysis. Mechanical Systems and Signal Processing. 2008; **22**(5): 1107–1132.

[8] Wu Z., Huang N.E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis. 2009; **1**(1): 1–41.

[9] Guo W., Huang L.J., Chen C., Zou H.W., Liu Z.W. Elimination of end effects in local mean decomposition using spectral coherence and applications for rotating machinery. Digital Signal Processing. 2016; **55**: 52–63.

[10] Lin D.C., Guo Z.L., An F.P., Zeng F.L. Elimination of end effects in empirical mode decomposition by mirror image coupled with support vector regression. Mechanical Systems and Signal Processing. 2012; **31**(8): 13–28.

[11] Antoni J. Cyclic spectral analysis in practice. Mechanical Systems and Signal Processing. 2007; **21**(2): 597–630.

[12] Wang D., Guo W., Tse P.W. An enhanced empirical mode decomposition method for blind component separation of a single-channel vibration signal mixture. Journal of Vibration & Control. 2015; **22**(11): 2603–2618.

[13] Guo W., Tse P.W. A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals. Journal of Sound and Vibration. 2013; **332**(2): 423–441.

[14] Ren D.Q., Yang S.X., Wu Z.T., Yan G.B. Research on end effect of LMD based time-frequency analysis in rotating machinery fault diagnosis. China Mechanical Engineering. 2012; **23**: 951–956.

[15] Guo W., Tse P.W. Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis. In: Proceedings of Prognostics & Health Management Conference 2010; 12–14 Jan. 2010; Macau. p. 1–6.

[16] Antoni J., Randall R.B. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. Mechanical Systems and Signal Processing. 2006; **20**(2): 308–331.

[17] Guo W., Tse P.W., Djordjevich A. Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. Measurement. 2012; **45**(5): 1308–1322.

[18] Antoni J. Fast computation of the kurtogram for the detection of transient faults. Mechanical Systems and Signal Processing. 2007; **21**(1): 108–124.

[19] Dybała J., Zimroz R. Rolling bearing diagnosing method based on Empirical Mode Decomposition of machine vibration signal. Applied Acoustics. 2014; **77**(3): 195–203.

[20] Guo W., Huang L.J., Chen C. Adaptive selection of feature signals in local mode decomposition for rotating machinery. In: Proceedings of The 7th Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modelling (APARM 2016); 24-26 August 2016; Seoul, Korea. 2016. p. 1–8.

[21] Georgoulas G., Loutas T., Stylios C.D., Kostopoulos V. Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. Mechanical Systems and Signal Processing. 2013; **41**(1–2): 510–525.

[22] Markou M., Singh S. Novelty detection: a review—part 1: statistical approaches. Signal Processing. 2003; **83**(12): 2481–2497.

[23] Polikar R. Ensemble based systems in decision making. IEEE Circuits & Systems Magazine. 2006; **6**(3): 21–45.

[24] Lei Y.G., He Z.J., Zi Y.Y. Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems with Applications. 2009; **36**(6): 9941–9948.

[25] Feldman M. Theoretical analysis and comparison of the Hilbert transform decomposition methods. Mechanical Systems and Signal Processing. 2008; **22**(3): 509–519.

[26] Cheng J.S. Yu D.J. Yang Y. Research on the intrinsic mode function (IMF) criterion in EMD method. Mechanical Systems & Signal Processing. 2006; **20**(4): 817–824.

[27] Rilling G., Flandrin P., Goncalves P. On empirical mode decomposition and its algorithms. In: Proceedings of IEEE EURASIP Workshop on Nonlinear Signal and Image Processing, Grado (I); June; 2003. p. 1–5.

[28] Delechelle E., Lemoine J., Niang O. Empirical mode decomposition: an analytical approach for sifting process. IEEE Signal Processing Letters. 2005; **12**(11): 764–767.

Chapter 3

Enhanced Principles in Design of Adaptive Fault Observers

Dušan Krokavec, Anna Filasová and Pavol Liščinský

Additional information is available at the end of the chapter

http://dx.doi.org/0.5772/67133

Abstract

In this chapter, modified techniques for fault estimation in linear dynamic systems are proposed, which give the possibility to simultaneously estimate the system state as well as slowly varying faults. Using the continuous-time adaptive observer form, the considered faults are assumed to be additive, thereby the principles can be applied for a broader class of fault signals. Enhanced algorithms using H_∞ approach are provided to verify stability of the observers, giving algorithms with improved performance of fault estimation. Exploiting the procedure for transforming the model with additive faults into an extended form, the proposed technique allows to obtain fault estimates that can be used for fault compensation in the fault tolerant control scheme. Analyzing the ambit of performances given on the mixed H_2/H_∞ design of the fault tolerant control, the joint design conditions are formulated as a minimization problem subject to convex constraints expressed by a system of linear matrix inequalities. Applied enhanced design conditions increase estimation rapidity also in noise environment and formulate a general framework for fault estimation using augmented or adaptive observer structures and active fault tolerant control in linear dynamic systems.

Keywords: linear dynamic systems, additive fault estimation, fault tolerant control design, enhanced bounded real lemma, linear matrix inequalities, H_∞ norm, H_2/H_∞ control strategy

1. Introduction

A model-based fault tolerant control (FTC) can be realized as control-laws set dependent, exploiting fault detection and isolation decision to reconfigure the control structure or as fault estimation dependent, preferring fault compensation within robust control framework. While integration of FTC with the fault localization decision technique requires a selection of optimal residual thresholds as well as a robust and stable reconfiguration mechanism [1], the fault

estimation-dependent FTC structures eliminate a threshold subjectivism and integrate FTC and estimation problems into one robust optimization task [2]. The realization is conditioned by observers, which performs the state reconstruction from the available signals.

The approach, in which faults estimates are used in a control structure to compensate the effects of acting faults, is adopted in modern FTC techniques [3, 4]. FTC with fault estimation for linear systems subject to bounded actuator or sensor faults, are proposed in [5]. The observer structures are in the Luenberger form [6] or realized as unknown input fault observers [7]. To guarantee the desired time response, a linear matrix inequality (LMI) based regional pole placement design strategy is proposed in [8] but such formulation introduces additive LMIs, which increase conservatism of the solutions. To minimize the set of LMIs of the circle regional pole placement is used; a modified approach in LMI construction is proposed in Ref. [9].

To estimate the actuator faults for the linear time-invariant systems without external disturbance the principles based on adaptive observers are frequently used, which make the estimation of the actuator faults by integrating the system output errors [10]. First introduced in Ref. [11], this principle was applied also for descriptor systems [7], linear systems with time delays [12], system with nonlinear dynamics [13], and a class of nonlinear systems described by Takagi-Sugeno fuzzy models [14, 15]. Some generalizations can be found in [16].

The H_2-norm is one of the most important characteristics of linear time-invariant control systems and so the problems concerning H_2, as well as H_∞, control have been studied by many authors (see, e.g. [17–20] and the references therein). Adding H_2 objective to H_∞ control design, a mixed H_2/H_∞ control problem was formulated in Ref. [21], with the goal to minimize H_2 norm subject to the constraint on H_∞ norm of the system transfer function. Such integrated design strategy corresponds to the optimization of the design parameters to satisfy desired specifications and to optimize the performance of the closed-loop system. Because of the importance of the control systems with these properties, considerable attention was dedicated to mixed H_2/H_∞ closed-loop performance criterion in design [22, 23] as well as to formulate the LMI-based computational technique [24, 25] to solve them or to exploit multiobjective algorithms for nonlinear, nonsmooth optimization in this design task [26, 27].

To guarantee suitable dynamics, new LMI conditions are proposed in the chapter for designing the fault observers as well as FTCs. Comparing with Ref. [5], the extended approach to the $\mathcal{D}$-stability introduced in Ref. [28] is used to minimize the number of LMIs in mixed H_2/H_∞ formulation of the FTC design and the eigenvalue circle clustering in fault observer design. In addition, different from Ref. [29], PD fault observer terms are comprehended through the enhanced descriptor approach [30], and a new design criterion is constructed in terms of LMIs. Since extended Lyapunov functions are exploited, the proposed approach offers the same degree of conservatism as the standard formulations [2, 31] but the H_∞ conditions are regularized under acting of H_2 constraint. Over and above, the $\mathcal{D}$-stability approach supports adjusting the fault estimator characteristics according to the fault frequency band.

The content and scope of the chapter are as follows. Placed after the introduction presented in Section 1, the basic preliminaries are given in Section 2. Section 3 reviews the definition and results concerning the adaptive fault observer design for continuous-time linear systems, Section 4 details the observer dynamic analysis and derives new results when using

the $\mathcal{D}$-stability circle criterion and Section 5 recasts the extended design conditions in the framework of LMIs based on structured matrix parameters. Then, in response to fault compensation principle for such type of fault observers, Section 6 derives the design conditions for the fault tolerant control structures, reflecting the joined H_2/H_∞ control idea. The relevance of the proposed approach is illustrated by a numerical example in Section 7 and Section 8 draws some concluding remarks.

2. Basic preliminaries

In order to analyze whether a linear MIMO system is stable under defined quadratic constraints, the basic properties can be summarized by the following LMI forms.

Considering linear MIMO systems

$$\dot{\boldsymbol{q}}(t) = \boldsymbol{A}\boldsymbol{q}(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{D}\boldsymbol{d}(t) \tag{1}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{q}(t) \tag{2}$$

where $\boldsymbol{q}(t)\in\mathbb{R}^n$, $\boldsymbol{u}(t)\in\mathbb{R}^r$, and $\boldsymbol{y}(t)\in\mathbb{R}^m$ are vectors of the system state, input, and output variables, respectively, $\boldsymbol{d}(t)\in\mathbb{R}^w$ is the unknown disturbance vector, $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ is the system dynamic matrix, $\boldsymbol{D}\in\mathbb{R}^{n\times w}$ is the disturbance input matrix, and $\boldsymbol{B}\in\mathbb{R}^{n\times r}$, $\boldsymbol{C}\in\mathbb{R}^{m\times n}$ are the system input and output matrices, then the system transfer functions matrices are

$$\boldsymbol{G}(s) = \boldsymbol{C}(s\boldsymbol{I}_n - \boldsymbol{A})^{-1}\boldsymbol{B}, \quad \boldsymbol{G}_d(s) = \boldsymbol{C}(s\boldsymbol{I}_n - \boldsymbol{A})^{-1}\boldsymbol{D} \tag{3}$$

where $\boldsymbol{I}_n\in\mathbb{R}^{n\times n}$ is an unitary matrix and the complex number s is the transform variable (Laplace variable) of the Laplace transform [32].

To characterize the system properties the following lemmas can be used.

Lemma 1 (*Lyapunov inequality*) [33] *The matrix $\boldsymbol{A}$ is Hurwitz if there exists a symmetric positive definite matrix $\boldsymbol{T}\in\mathbb{R}^{n\times n}$ such that*

$$\boldsymbol{T} = \boldsymbol{T}^T > 0, \quad \boldsymbol{A}^T\boldsymbol{T} + \boldsymbol{T}\boldsymbol{A} < 0 \tag{4}$$

Lemma 2 [34] *The matrix $\boldsymbol{A}$ is Hurwitz and $\|\boldsymbol{G}(s)\|_2 < \gamma_2$ if there exists a symmetric positive definite matrix $\boldsymbol{V}\in\mathbb{R}^{n\times n}$ and a positive scalar $\gamma_2\in\mathbb{R}$, such that*

$$\boldsymbol{V} = \boldsymbol{V}^T > 0 \tag{5}$$

$$\boldsymbol{A}\boldsymbol{V} + \boldsymbol{V}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{B}^T < 0 \tag{6}$$

$$tr(\boldsymbol{C}\boldsymbol{V}\boldsymbol{C}^T) < \gamma_2^2 \tag{7}$$

where $\gamma_2 > 0$, $\gamma_2\in\mathbb{R}$ is H_2 norm of the transfer function matrix $\boldsymbol{G}(s)$.

Lemma 3 (*Bounded real lemma*) [35] *The matrix $\boldsymbol{A}$ is Hurwitz and $\|\boldsymbol{G}_d(s)\|_\infty < \gamma_\infty$ if there exists a symmetric positive definite matrix $\boldsymbol{U}\in\mathbb{R}^{n\times n}$ and a positive scalar $\gamma_\infty\in\mathbb{R}$ such that*

$$\boldsymbol{U} = \boldsymbol{U}^T > 0 \tag{8}$$

$$\begin{bmatrix} \boldsymbol{U}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{U} & * & * \\ \boldsymbol{D}^T\boldsymbol{U} & -\gamma_\infty \boldsymbol{I}_w & * \\ \boldsymbol{C} & \boldsymbol{0} & -\gamma_\infty \boldsymbol{I}_m \end{bmatrix} < 0, \tag{9}$$

where $\boldsymbol{I}_w \in \mathbb{R}^{w\times w}$, $\boldsymbol{I}_m \in \mathbb{R}^{m\times m}$ *are identity matrices and* $\gamma_\infty > 0$, $\gamma_\infty \in \mathbb{R}$ *is* H_∞ *norm of the disturbance transfer function matrix* $\boldsymbol{G}_d(s)$.

*Hereafter, * denotes the symmetric item in a symmetric matrix.*

Lemma 4 [28] *The matrix* **A** *is* $\mathcal{D}$*-stable Hurwitz if for given positive scalars* $a, \varrho \in \mathbb{R}$, $a > \varrho$, *there exists a symmetric positive definite matrix* $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ *such that*

$$\boldsymbol{T} = \boldsymbol{T}^T > 0, \tag{10}$$

$$\begin{bmatrix} -\varrho\boldsymbol{T} & * \\ \boldsymbol{T}\boldsymbol{A} + a\boldsymbol{T} & -\varrho\boldsymbol{T} \end{bmatrix} < 0, \tag{11}$$

while the eigenvalues of $\boldsymbol{A}$ *are clustered in the circle with the origin* $c_o = (-a + 0i)$ *and radius* ϱ *within the complex plane S.*

Lemma 5 (*Schur complement*) [36] *Let* $\boldsymbol{O}$ *be a real matrix, and* $\boldsymbol{N}$ ($\mathbf{M}$) *be a positive definite symmetric matrix of appropriate dimension, then the following inequalities are equivalent*

$$\begin{bmatrix} \boldsymbol{M} & \boldsymbol{O} \\ \boldsymbol{O}^T & -\boldsymbol{N} \end{bmatrix} < 0 \Leftrightarrow \begin{bmatrix} \boldsymbol{M} + \boldsymbol{O}\boldsymbol{N}^{-1}\boldsymbol{O}^T & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{N} \end{bmatrix} < 0 \Leftrightarrow \boldsymbol{M} + \boldsymbol{O}\boldsymbol{N}^{-1}\boldsymbol{O}^T < 0, \quad \boldsymbol{N} > 0, \tag{12}$$

$$\begin{bmatrix} -\boldsymbol{M} & \boldsymbol{O} \\ \boldsymbol{O}^T & \boldsymbol{N} \end{bmatrix} < 0 \Leftrightarrow \begin{bmatrix} -\boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N} + \boldsymbol{O}^T\boldsymbol{M}^{-1}\boldsymbol{O} \end{bmatrix} < 0 \Leftrightarrow \boldsymbol{N} + \boldsymbol{O}^T\boldsymbol{M}^{-1}\boldsymbol{O} < 0, \quad \boldsymbol{M} > 0. \tag{13}$$

Lemma 6 (*Krasovskii lemma*) [37] *The autonomous system (1) is asymptotically stable if for a given symmetric positive semidefinite matrix* $\boldsymbol{L} \in \mathbb{R}^{n\times n}$ *there exists a symmetric positive definite matrix* $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ *such that*

$$\boldsymbol{T} = \boldsymbol{T}^T > 0, \tag{14}$$

$$\boldsymbol{A}^T\boldsymbol{T} + \boldsymbol{T}\boldsymbol{A} + \boldsymbol{L} < 0, \tag{15}$$

where $\boldsymbol{L}$ *is the weight matrix of an integral quadratic constraint interposed on the state vector* $\boldsymbol{q}(t)$.

3. Proportional adaptive fault observers

To characterize the role of constraints in the proposed methodology and ease of understanding the presented approach, the theorems' proofs are restated in a condensed form in this section and also for theorems already being presented by the authors, e.g., in Refs. [38–40].

Despite different definitions, the best description for the formulation of the problem is based on the common state-space description of the linear dynamic multiinput, multioutput (MIMO) systems in the presence of unknown faults of the form

$$\dot{q}(t) = \boldsymbol{A}\boldsymbol{q}(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{F}\boldsymbol{f}(t), \tag{16}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{q}(t), \tag{17}$$

where $\boldsymbol{q}(t)\in\mathbb{R}^n$, $\boldsymbol{u}(t)\in\mathbb{R}^r$, and $\boldsymbol{y}(t)\in\mathbb{R}^m$ are vectors of the system, input, and output variables, respectively, $\boldsymbol{f}(t)\in\mathbb{R}^p$ is the unknown fault vector, $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ is the system dynamics matrix, $\boldsymbol{F}\in\mathbb{R}^{n\times p}$ is the fault input matrix, and $\boldsymbol{B}\in\mathbb{R}^{n\times r}$ and $\boldsymbol{C}\in\mathbb{R}^{m\times n}$ are the system input and output matrices, $m, r, p < n$,

$$\operatorname{rank}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{F}\\ \boldsymbol{C} & \boldsymbol{0}\end{bmatrix} = n + p, \tag{18}$$

and the couple ($\boldsymbol{A}$,$\boldsymbol{C}$) is observable.

Limiting to the time-invariant system (16) and (17) to estimate the faults and the system states simultaneously, as well as focusing on slowly varying additive faults, the adaptive fault observer is considered in the following form [41]

$$\dot{\boldsymbol{q}}_e(t) = \boldsymbol{A}\boldsymbol{q}_e(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{F}\boldsymbol{f}_e(t) + \boldsymbol{J}(\boldsymbol{y}(t)-\boldsymbol{y}_e(t)), \tag{19}$$

$$\boldsymbol{y}_e(t) = \boldsymbol{C}\boldsymbol{q}_e(t), \tag{20}$$

where $\boldsymbol{q}_e(t)\in\mathbb{R}^n$, $\boldsymbol{y}_e(t)\in\mathbb{R}^m$, and $\boldsymbol{f}_e(t)\in\mathbb{R}^p$ are estimates of the system states vector, the output variables vector, and the fault vector, respectively, and $\boldsymbol{J}\in\mathbb{R}^{n\times m}$ is the observer gain matrix.

The observer (19) and (20) is combined with the fault estimation updating law of the form [42]

$$\dot{\boldsymbol{f}}_e(t) = \boldsymbol{G}\boldsymbol{H}^T\boldsymbol{e}_y(t), \quad \boldsymbol{e}_y(t) = \boldsymbol{y}(t)-\boldsymbol{y}_e(t) = \boldsymbol{C}\boldsymbol{e}_q(t), \quad \boldsymbol{e}_q(t) = \boldsymbol{q}(t)-\boldsymbol{q}_e(t), \tag{21}$$

where $\boldsymbol{H}\in\mathbb{R}^{m\times p}$ is the gain matrix and $\boldsymbol{G} = \boldsymbol{G}^T > 0$, $\boldsymbol{G}\in\mathbb{R}^{p\times p}$ is a learning weight matrix that has to be set interactively in the design step.

In order to express unexpectedly changing faults as a function of the system and observer outputs and to apply the adaptive estimation principle, it is considered that the fault vector is piecewise constant, differentiable, and bounded, i.e., $\|\boldsymbol{f}(t)\|\leq\boldsymbol{f}_{\max} < \infty$, the upper bound norm $\boldsymbol{f}_{\max}$ is known, and the value of $\boldsymbol{f}(t)$ is set to zero vector until a fault occurs. This assumption, in general, implies that the time derivative of $\boldsymbol{e}_f(t)$ can be considered as

$$\dot{\boldsymbol{f}}(t)\approx\boldsymbol{0}, \quad \dot{e}_f(t) = -\dot{\boldsymbol{f}}_e(t), \quad \boldsymbol{e}_f(t) = \boldsymbol{f}(t)-\boldsymbol{f}_e(t). \tag{22}$$

These assumptions have to be taking into account by designing the matrix parameters of the observers to ensure asymptotic convergence of the estimation errors, Eqs. (21) and (22). The task is to design the matrix $\boldsymbol{J}$ in such a way that the observer dynamics matrix $\boldsymbol{A}_e = \boldsymbol{A}-\boldsymbol{J}\boldsymbol{C}$ is stable and $\boldsymbol{f}_e(t)$ approximates a slowly varying actuator fault $\boldsymbol{f}(t)$.

3.1. Design conditions

If single faults influence the system through different input vectors (columns of the matrix $\boldsymbol{F}$), it is possible to avoid designing the estimators with the tuning matrix parameter $\boldsymbol{G} > 0$ and formulate the design task through the set of LMIs and a linear matrix equality.

Theorem 1 *The adaptive fault observer (19) and (20) is stable if there exists a symmetric positive definite matrix* $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ *and matrices* $\boldsymbol{H} \in \mathbb{R}^{n \times p}$, $\boldsymbol{Y} \in \mathbb{R}^{n \times m}$ *such that*

$$\boldsymbol{P} = \boldsymbol{P}^T > 0, \tag{23}$$

$$\boldsymbol{PA} + \boldsymbol{A}^T\boldsymbol{P} - \boldsymbol{YC} - \boldsymbol{C}^T\boldsymbol{Y}^T < 0, \tag{24}$$

$$\boldsymbol{PF} = \boldsymbol{C}^T\boldsymbol{H}. \tag{25}$$

When the above conditions hold, the observer gain matrix is given by

$$\boldsymbol{J} = \boldsymbol{P}^{-1}\boldsymbol{Y} \tag{26}$$

and the adaptive fault estimation algorithm is

$$\dot{\boldsymbol{f}}_e(t) = \boldsymbol{GH}^T\boldsymbol{Ce}_q(t), \tag{27}$$

where

$$\boldsymbol{e}_q(t) = \boldsymbol{q}(t) - \boldsymbol{q}_e(t) \tag{28}$$

and $\boldsymbol{G} \in \mathbb{R}^{p \times p}$ *is a symmetric positive definite matrix which values are set interactive in design.*

Proof. From the system models (16) and (17) and the observer models (19) and (20), it can be obtained that

$$\begin{aligned} \dot{\boldsymbol{e}}_q(t) &= \boldsymbol{Aq}(t) + \boldsymbol{Bu}(t) + \boldsymbol{Ff}(t) - \boldsymbol{Aq}_e(t) - \boldsymbol{Bu}(t) - \boldsymbol{Ff}_e(t) - \boldsymbol{J}(\boldsymbol{y}(t) - \boldsymbol{y}_e(t)) = \\ &= (\boldsymbol{A} - \boldsymbol{JC})\boldsymbol{e}_q(t) + \boldsymbol{Fe}_f(t) = \boldsymbol{A}_e\boldsymbol{e}_q(t) + \boldsymbol{Fe}_f(t), \end{aligned} \tag{29}$$

where the observer system matrix is

$$\boldsymbol{A}_e = \boldsymbol{A} - \boldsymbol{JC}. \tag{30}$$

Since $\boldsymbol{e}_q(t)$ is linear with respect to the system parameters, it is possible to consider the Lyapunov function candidate in the following form

$$v(\boldsymbol{e}_q(t)) = \boldsymbol{e}_q^T(t)\boldsymbol{Pe}_q(t) + \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) > 0, \tag{31}$$

where $\boldsymbol{P}$, $\boldsymbol{G}$ are real, symmetric, and positive definite matrices. Then, the time derivative of $v(\boldsymbol{e}_q(t))$ is

$$\dot{v}(\boldsymbol{e}_q(t)) = \dot{v}_0(\boldsymbol{e}_q(t)) + \dot{v}_1(\boldsymbol{e}_q(t)) < 0, \tag{32}$$

where

$$\begin{aligned}\dot{v}_0(\boldsymbol{e}_q(t)) &= \dot{e}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{e}_q(t) = \\ &= (\boldsymbol{A}_e\boldsymbol{e}_q(t) + \boldsymbol{F}\boldsymbol{e}_f(t))^T\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}(\boldsymbol{A}_e\boldsymbol{e}_q(t) + \boldsymbol{F}\boldsymbol{e}_f(t)) = \\ &= \boldsymbol{e}_q^T(t)(\boldsymbol{A}_e^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}_e)\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\boldsymbol{F}\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)\boldsymbol{F}^T\boldsymbol{P}\boldsymbol{e}_q(t),\end{aligned} \tag{33}$$

$$\dot{v}_1(\boldsymbol{e}_q(t)) = \dot{e}_f^T(t)\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\dot{e}_f(t) = -\boldsymbol{f}_e^T(t)\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) - \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\dot{\boldsymbol{f}}_e(t). \tag{34}$$

Inserting Eq. (21) into Eq. (34) leads to

$$\begin{aligned}\dot{v}_1(\boldsymbol{e}_q(t)) &= -\boldsymbol{e}_q^T(t)\boldsymbol{C}^T\boldsymbol{H}\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) - \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\boldsymbol{G}\boldsymbol{H}^T\boldsymbol{C}\boldsymbol{e}_q(t) \\ &= -\boldsymbol{e}_q^T(t)\boldsymbol{C}^T\boldsymbol{H}\boldsymbol{e}_f(t) - \boldsymbol{e}_f^T(t)\boldsymbol{H}^T\boldsymbol{C}\boldsymbol{e}_q(t)\end{aligned} \tag{35}$$

and substituting Eq. (35) with Eq. (30) into Eq. (33), the following inequality is obtained

$$\begin{aligned}\dot{v}(\boldsymbol{e}_q(t)) &= \boldsymbol{e}_q^T(t)\Big((\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})^T\boldsymbol{P} + \boldsymbol{P}(\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})\Big)\boldsymbol{e}_q(t) \\ &+\boldsymbol{e}_q^T(t)(\boldsymbol{P}\boldsymbol{F}-\boldsymbol{C}^T\boldsymbol{H})\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)(\boldsymbol{F}^T\boldsymbol{P}-\boldsymbol{H}^T\boldsymbol{C})\boldsymbol{e}_q(t) < 0.\end{aligned} \tag{36}$$

It is clear that the requirement

$$\boldsymbol{e}_q^T(t)(\boldsymbol{P}\boldsymbol{F}-\boldsymbol{C}^T\boldsymbol{H})\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)(\boldsymbol{F}^T\boldsymbol{P}-\boldsymbol{H}^T\boldsymbol{C})\boldsymbol{e}_q(t) = 0 \tag{37}$$

can be satisfied when Eq. (25) is satisfied.

Using the above given condition (37), the resulting formula for $\dot{v}(\boldsymbol{e}_q(t))$ takes the form

$$\dot{v}(\boldsymbol{e}_q(t)) = \boldsymbol{e}_q^T(t)((\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})^T\boldsymbol{P} + \boldsymbol{P}(\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C}))\boldsymbol{e}_q(t) < 0, \tag{38}$$

and the LMI, defining the observer stability condition, is presented as

$$\boldsymbol{P}(\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C}) + (\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})^T\boldsymbol{P} < 0. \tag{39}$$

Introducing the notation

$$\boldsymbol{P}\boldsymbol{J} = \boldsymbol{Y} \tag{40}$$

it is possible to express Eq. (39) as Eq. (24). This concludes the proof.

3.2. Enhanced design conditions

The observer stability analysis could be carried out generally under the assumption (29), i.e., using the forced differential equation of the form

$$\dot{\boldsymbol{e}}_q(t) = (\boldsymbol{A}-\boldsymbol{JC})\boldsymbol{e}_q(t) + \boldsymbol{F}\boldsymbol{e}_f(t), \tag{41}$$

$$\boldsymbol{e}_y(t) = \boldsymbol{C}\boldsymbol{e}_q(t), \tag{42}$$

while

$$\boldsymbol{G}_f(s) = \boldsymbol{C}(\boldsymbol{A}-\boldsymbol{JC})^{-1}\boldsymbol{F}. \tag{43}$$

It is evident now that $\boldsymbol{e}_f(t)$ acts on the state error dynamics as an unknown disturbance and, evidently, this differential equation is so not autonomous after a fault occurrence. Reflecting this fact, the enhanced approach is proposed to decouple Lyapunov matrix $\boldsymbol{P}$ from the system matrices $\boldsymbol{A}$, $\boldsymbol{C}$ by introducing a slack matrix $\boldsymbol{Q}$ in the observer stability condition, as well as to decouple the tuning parameter δ from the matrix $\boldsymbol{G}$ in the learning rate setting and using δ to tune the observer dynamic properties. Since the design principle for unknown input observer cannot be used, the impact of faults on observer dynamics is moreover minimized with respect to the H_∞ norm of the transfer functions matrix of $\boldsymbol{G}_f(s)$, while a reduction in the fault amplitude estimate is easily countervailing using the matrix $\boldsymbol{G}$. In this sense the enhanced design conditions can be formulated in the following way.

Theorem 2 *The adaptive fault observer (19) and (20) is stable if for a given positive $\delta\in\mathbb{R}$ there exist symmetric positive definite matrices $\boldsymbol{P}\in\mathbb{R}^{n\times n}$, $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$, matrices $\boldsymbol{H}\in\mathbb{R}^{n\times p}$, $\boldsymbol{Y}\in\mathbb{R}^{n\times m}$ and a positive scalar $\gamma\in\mathbb{R}$ such that*

$$\boldsymbol{P} = \boldsymbol{P}^T > 0, \quad \boldsymbol{Q} = \boldsymbol{Q}^T > 0, \quad \gamma > 0, \tag{44}$$

$$\begin{bmatrix} \boldsymbol{QA} + \boldsymbol{A}^T\boldsymbol{Q} - \boldsymbol{YC} - \boldsymbol{C}^T\boldsymbol{Y}^T & * & * & * \\ \boldsymbol{P} - \boldsymbol{Q} + \delta\boldsymbol{QA} - \delta\boldsymbol{YC} & -2\delta\boldsymbol{Q} & * & * \\ \boldsymbol{0} & \delta\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & * \\ \boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0, \tag{45}$$

$$\boldsymbol{QF} = \boldsymbol{C}^T\boldsymbol{H}. \tag{46}$$

When the above conditions are affirmative the estimator gain matrix is given by the relation

$$\boldsymbol{J} = \boldsymbol{Q}^{-1}\boldsymbol{Y}. \tag{47}$$

Proof. Using Krasovskii lemma, the Lyapunov function candidate can be considered as

$$v(\boldsymbol{e}_q(t)) = \boldsymbol{e}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) + \gamma^{-1}\int_0^t (\boldsymbol{e}_y^T(r)\boldsymbol{e}_y(r) - \gamma^2\boldsymbol{e}_f^T(r)\boldsymbol{e}_f(r))dr > 0, \tag{48}$$

where $\boldsymbol{P} = \boldsymbol{P}^T > 0$, $\boldsymbol{G} = \boldsymbol{G}^T > 0$, $\gamma > 0$, and γ is an upper bound of H_∞ norm of the transfer function matrix $\boldsymbol{G}_f(s)$. Then the time derivative of $v(\boldsymbol{e}_q(t))$ has to be negative, i.e.,

$$\dot{v}(\boldsymbol{e}_q(t)) = \dot{e}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{\boldsymbol{e}}_q(t) + \dot{\boldsymbol{e}}_f^T(t)\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\dot{\boldsymbol{e}}_f(t) + \gamma^{-1}\boldsymbol{e}_y^T(t)\boldsymbol{e}_y(t) - \gamma\boldsymbol{e}_f^T(t)\boldsymbol{e}_f(t) < 0. \tag{49}$$

If it is assumed that Eqs. (34) and (35) hold, then the substitution of Eq. (35) into Eq. (49) leads to

$$\dot{v}(\boldsymbol{e}_q(t)) = \dot{e}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{\boldsymbol{e}}_q(t) - \boldsymbol{e}_q^T(t)\boldsymbol{C}^T\boldsymbol{H}\boldsymbol{e}_f(t) - \boldsymbol{e}_f^T(t)\boldsymbol{H}^T\boldsymbol{C}\boldsymbol{e}_q(t) + \gamma^{-1}\boldsymbol{e}_y^T(t)\boldsymbol{e}_y(t) - \gamma\boldsymbol{e}_f^T(t)\boldsymbol{e}_f(t) < 0. \tag{50}$$

Since Eq. (41) implies

$$(\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})\boldsymbol{e}_q(t) + \boldsymbol{F}\boldsymbol{e}_f(t) - \dot{\boldsymbol{e}}_q(t) = 0, \tag{51}$$

it is possible to define the following condition based on the equality (51)

$$(\boldsymbol{e}_q^T(t)\boldsymbol{Q} + \dot{e}_q^T(t)\delta\boldsymbol{Q})((\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})\boldsymbol{e}_q(t) + \boldsymbol{F}\boldsymbol{e}_f(t) - \dot{\boldsymbol{e}}_q(t)) = 0, \tag{52}$$

where $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$ is a symmetric positive definite matrix and $\delta\in IR$ is a positive scalar.

Then, adding Eq. (52) and its transposition to Eq. (50), the following has to be satisfied

$$\begin{aligned}\dot{v}(\boldsymbol{e}_q(t)) = {} & \dot{e}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{\boldsymbol{e}}_q(t) - \boldsymbol{e}_q^T(t)\boldsymbol{C}^T\boldsymbol{H}\boldsymbol{e}_f(t) - \boldsymbol{e}_f^T(t)\boldsymbol{H}^T\boldsymbol{C}\boldsymbol{e}_q(t) \\ & + (\boldsymbol{e}_q^T(t)\boldsymbol{Q} + \dot{e}_q^T(t)\delta\boldsymbol{Q})((\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})\boldsymbol{e}_q(t) - \dot{\boldsymbol{e}}_q(t)) + \gamma^{-1}\boldsymbol{e}_y^T(t)\boldsymbol{e}_y(t) \\ & + ((\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})\boldsymbol{e}_q(t) - \dot{\boldsymbol{e}}_q(t))^T(\boldsymbol{Q}\boldsymbol{e}_q(t) + \delta\boldsymbol{Q}\dot{\boldsymbol{e}}_q(t)) - \gamma\boldsymbol{e}_f^T(t)\boldsymbol{e}_f(t) \\ & + (\boldsymbol{e}_q^T(t)\boldsymbol{Q} + \dot{e}_q^T(t)\delta\boldsymbol{Q})\boldsymbol{F}\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)\boldsymbol{F}^T(\boldsymbol{Q}\boldsymbol{e}_q(t) + \delta\boldsymbol{Q}\dot{\boldsymbol{e}}_q(t)) < 0.\end{aligned} \tag{53}$$

If the following requirement is introduced

$$\boldsymbol{e}_f^T(t)(\boldsymbol{F}^T\boldsymbol{Q} - \boldsymbol{H}^T\boldsymbol{C})\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)(\boldsymbol{Q}\boldsymbol{F} - \boldsymbol{C}^T\boldsymbol{H})\boldsymbol{e}_f(t) = 0, \tag{54}$$

it is obvious that Eq. (54) can be satisfied when Eq. (46) is satisfied. Thus, the condition (54) allows to write Eq. (53) as follows

$$\begin{aligned}\dot{v}(\boldsymbol{e}_q(t)) = {} & \dot{e}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{\boldsymbol{e}}_q(t) + \boldsymbol{e}_q^T(t)\gamma^{-1}\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{e}_q(t) - \gamma\boldsymbol{e}_f^T(t)\boldsymbol{e}_f(t) + \\ & + (\boldsymbol{e}_q^T(t)\boldsymbol{Q} + \dot{e}_q^T(t)\delta\boldsymbol{Q})((\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})\boldsymbol{e}_q(t) - \dot{\boldsymbol{e}}_q(t)) + \\ & + (\boldsymbol{e}_q^T(t)(\boldsymbol{A}-\boldsymbol{J}\boldsymbol{C})^T - \dot{\boldsymbol{e}}_q^T(t))(\boldsymbol{Q}\boldsymbol{e}_q(t) + \delta\boldsymbol{Q}\dot{\boldsymbol{e}}_q(t)) + \\ & + \dot{\boldsymbol{e}}_q^T(t)\delta\boldsymbol{Q}\boldsymbol{F}\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)\delta\boldsymbol{F}^T\boldsymbol{Q}\dot{\boldsymbol{e}}_q(t) < 0.\end{aligned} \tag{55}$$

Relying on Eq. (55), it is possible to write the observer stability condition as

$$\dot{v}(\boldsymbol{e}_d(t)) = \boldsymbol{e}_d^T(t)\boldsymbol{P}_d\boldsymbol{e}_d(t) < 0, \tag{56}$$

where the following notations

$$\boldsymbol{P}_d = \begin{bmatrix} \boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) + (\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} + \gamma^{-1}\boldsymbol{C}^T\boldsymbol{C} & \boldsymbol{P}-\boldsymbol{Q} + \delta(\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{P}-\boldsymbol{Q} + \delta\boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) & -2\delta\boldsymbol{Q} & \delta\boldsymbol{QF} \\ \boldsymbol{0} & \delta\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p \end{bmatrix} < 0, \tag{57}$$

$$\boldsymbol{e}_d^T(t) = [\,\boldsymbol{e}_q^T(t) \quad \dot{\boldsymbol{e}}_q^T(t)\ \boldsymbol{e}_f^T(t)], \tag{58}$$

are exploited.

Introducing the substitution

$$\boldsymbol{QJ} = \boldsymbol{Y} \tag{59}$$

and using the Schur complement property with respect to the item $\gamma^{-1}\boldsymbol{C}^T\boldsymbol{C}$, then Eq. (57) implies Eq. (45). This concludes the proof.

Theorem 3 *The adaptive fault observer (19) and (20) is stable if there exists a symmetric positive definite matrix* $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$, *matrices* $\boldsymbol{H}\in\mathbb{R}^{n\times p}$, $\boldsymbol{Y}\in\mathbb{R}^{n\times m}$ *and a positive scalar* $\gamma\in\mathbb{R}$ *such that*

$$\boldsymbol{Q} = \boldsymbol{Q}^T > 0, \quad \gamma > 0, \tag{60}$$

$$\begin{bmatrix} \boldsymbol{QA} + \boldsymbol{A}^T\boldsymbol{Q} - \boldsymbol{YC} - \boldsymbol{C}^T\boldsymbol{Y}^T & * & * \\ \boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & * \\ \boldsymbol{C} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0. \tag{61}$$

$$\boldsymbol{QF} = \boldsymbol{C}^T\boldsymbol{H}. \tag{62}$$

When the above conditions are affirmative the estimator gain matrix is given by the relation

$$\boldsymbol{J} = \boldsymbol{Q}^{-1}\boldsymbol{Y}. \tag{63}$$

Proof. Premultiplying the left side and postmultiplying the right side of Eq. (57) by the transformation matrix

$$\boldsymbol{T}_x = \mathrm{diag}[\,\boldsymbol{I}_n \quad \delta^{-1}\boldsymbol{I}_n \quad \boldsymbol{I}_p \quad \boldsymbol{I}_m\,] \tag{64}$$

gives

$$\begin{bmatrix} \boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) + (\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} + \gamma^{-1}\boldsymbol{C}^T\boldsymbol{C} & \delta^{-1}(\boldsymbol{P}-\boldsymbol{Q}) + (\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} & \boldsymbol{0} \\ \delta^{-1}(\boldsymbol{P}-\boldsymbol{Q}) + \boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) & -2\delta^{-1}\boldsymbol{Q} & \boldsymbol{QF} \\ \boldsymbol{0} & \boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p \end{bmatrix} < 0. \tag{65}$$

Considering that $\boldsymbol{P} = \boldsymbol{Q}$ and using the Schur complement property, then the inequality (65) can be rewritten as

$$\begin{aligned} &\boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) + (\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} + \gamma^{-1}\boldsymbol{C}^T\boldsymbol{C} \\ &+(\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q}\frac{1}{2}\delta\boldsymbol{Q}^{-1}\boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) + [\,\boldsymbol{0} \atop \boldsymbol{QF}\,]\gamma^{-1}\boldsymbol{I}_p[\,\boldsymbol{0} \quad \boldsymbol{F}^T\boldsymbol{Q}\,] < 0. \end{aligned} \tag{66}$$

Since the first matrix element in the second row of Eq. (66) is zero matrix if $\delta = 0$ and considering that nonzero component unit of the last matrix element in this raw is certainly positive semidefinite, it can claim that

$$\boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) + (\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} + \gamma^{-1}\boldsymbol{C}^T\boldsymbol{C} + \boldsymbol{QF}\gamma^{-1}\boldsymbol{I}_p\boldsymbol{F}^T\boldsymbol{Q} < 0. \tag{67}$$

Thus, applying the Schur complement property, it can be written as

$$\begin{bmatrix} \boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) + (\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} + \gamma^{-1}\boldsymbol{C}^T\boldsymbol{C} & \boldsymbol{QF} \\ \boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p \end{bmatrix} < 0, \tag{68}$$

$$\begin{bmatrix} \boldsymbol{Q}(\boldsymbol{A}-\boldsymbol{JC}) + (\boldsymbol{A}-\boldsymbol{JC})^T\boldsymbol{Q} & \boldsymbol{QF} & \boldsymbol{C}^T \\ \boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0, \tag{69}$$

respectively. With the notation (59) then Eq. (69) gives Eqs. (61). This concludes the proof.

Comparing with Lemma 3, it can be seen that Eqs. (60)–(62) is an extended form of the bounded real lemma (BRL) structure, applicable in the design of proportional adaptive fault observers.

4. Observer dynamics with eigenvalues clustering in D-stability circle

Generalizing the approach covering decoupling of Lyapunov matrix from the observer system matrix parameters by using a slack matrix, with a good exposition of the given theorems, the observer eigenvalues placement in a circular $\mathcal{D}$-stability region is proposed to enable wide adaptation to faults dynamics.

Theorem 4 *The adaptive fault observer (19) and (20) is $\mathcal{D}$-stable if for given positive scalars $\delta, a, \varrho \in \mathbb{R}$, $a > \varrho$, there exist symmetric positive definite matrices $\boldsymbol{P} \in \mathbb{R}^{n\times n}$, $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, matrices $\boldsymbol{H} \in \mathbb{R}^{n\times p}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times m}$ and a positive scalar $\gamma \in \mathbb{R}$ such that*

$$\boldsymbol{P} = \boldsymbol{P}^T > 0, \quad \boldsymbol{Q} = \boldsymbol{Q}^T > 0, \quad \gamma > 0, \tag{70}$$

$$\begin{bmatrix} -\varrho\boldsymbol{Q} & * & * & * & * \\ a\boldsymbol{Q} + \boldsymbol{QA} - \boldsymbol{YC} & -\varrho\boldsymbol{Q} & * & * & * \\ \boldsymbol{P} - \boldsymbol{Q} + \frac{\delta}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} + \frac{\delta}{\varrho}\boldsymbol{QA} - \frac{\delta}{\varrho}\boldsymbol{YC} & \boldsymbol{0} & -2\delta\boldsymbol{Q} & * & * \\ \boldsymbol{0} & \boldsymbol{0} & \frac{\delta}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & * \\ \boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0, \tag{71}$$

$$\boldsymbol{QF} = \boldsymbol{C}^T\boldsymbol{H}. \tag{72}$$

When the above conditions are affirmative the estimator gain matrix can be computed as

$$\boldsymbol{J} = \boldsymbol{Q}^{-1}\boldsymbol{Y} \tag{73}$$

and the adaptive fault estimation algorithm is given by (27).

Proof. Choosing the Lyapunov function candidate as

$$\begin{aligned} v(\boldsymbol{e}_q(t)) = \boldsymbol{e}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) + \gamma^{-1}\int_0^t (\boldsymbol{e}_y^T(r)\boldsymbol{e}_y(r) - \gamma^2\boldsymbol{e}_f^T(r)\boldsymbol{e}_f(r))dr \\ + \varrho^{-1}\int_0^t \boldsymbol{e}_q^T(r)\boldsymbol{A}_e^T\boldsymbol{Q}\boldsymbol{A}_e\boldsymbol{e}_q(r)dr > 0, \end{aligned} \tag{74}$$

where $\boldsymbol{P} = \boldsymbol{P}^T > 0$, $\boldsymbol{G} = \boldsymbol{G}^T > 0$, $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$, $\gamma > 0$, γ is an upper bound of H_∞ norm of the transfer function matrix (43) and where the generalized observer differential equation takes the form [28]

$$\dot{e}_q(t) = \boldsymbol{A}_{er}\boldsymbol{e}_q(t) + \boldsymbol{F}_r\boldsymbol{e}_f(t), \tag{75}$$

while, with $a > 0$, $\varrho > 0$ such that $\varrho < a$, the matrices $\boldsymbol{A}_{cr}$, $\boldsymbol{F}_{rr}$ are given as

$$\boldsymbol{A}_{er} = \frac{a}{\varrho}\boldsymbol{A}_e + \frac{a^2 - \varrho^2}{2\varrho}\boldsymbol{I}_n, \quad \boldsymbol{F}_r = \frac{1}{\varrho}\boldsymbol{F}. \tag{76}$$

Then, the time derivative of $v(\boldsymbol{e}_q(t))$ is

$$\begin{aligned} \dot{v}(\boldsymbol{e}_q(t)) = \dot{\boldsymbol{e}}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{\boldsymbol{e}}_q(t) + \dot{\boldsymbol{e}}_f^T(t)\boldsymbol{G}^{-1}\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)\boldsymbol{G}^{-1}\dot{\boldsymbol{e}}_f(t) + \\ + \boldsymbol{e}_q^T(t)\boldsymbol{A}_e^T\varrho^{-1}\boldsymbol{Q}\boldsymbol{A}_e\boldsymbol{e}_q(t) + \gamma^{-1}\boldsymbol{e}_y^T(t)\boldsymbol{e}_y(t) - \gamma\boldsymbol{e}_f^T(t)\boldsymbol{e}_f(t) < 0. \end{aligned} \tag{77}$$

Assuming that, with respect to Eqs. (34) and (35), the inequality (50) holds, then Eq. (77) gives

$$\begin{aligned} \dot{v}(\boldsymbol{e}_q(t)) = \dot{\boldsymbol{e}}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{\boldsymbol{e}}_q(t) - \boldsymbol{e}_q^T(t)\boldsymbol{C}^T\boldsymbol{H}\boldsymbol{e}_f(t) - \boldsymbol{e}_f^T(t)\boldsymbol{H}^T\boldsymbol{C}\boldsymbol{e}_q(t) \\ + \boldsymbol{e}_q^T(t)\boldsymbol{A}_e^T\varrho^{-1}\boldsymbol{Q}\boldsymbol{A}_e\boldsymbol{e}_q(t) + \gamma^{-1}\boldsymbol{e}_y^T(t)\boldsymbol{e}_y(t) - \gamma\boldsymbol{e}_f^T(t)\boldsymbol{e}_f(t) < 0. \end{aligned} \tag{78}$$

Generalizing the equation (75), the following condition can be set

$$(\boldsymbol{e}_q^T(t)\boldsymbol{Q} + \dot{\boldsymbol{e}}_q^T(t)\delta\boldsymbol{Q})(\boldsymbol{A}_{er}\boldsymbol{e}_q(t) + \boldsymbol{F}_r\boldsymbol{e}_f(t) - \dot{\boldsymbol{e}}_q(t)) = 0, \tag{79}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix and $\delta \in IR$ is a positive scalar. Therefore, adding Eq. (79) and its transposition to Eq. (78) gives

$$\begin{aligned} \dot{v}(\boldsymbol{e}_q(t)) = \dot{\boldsymbol{e}}_q^T(t)\boldsymbol{P}\boldsymbol{e}_q(t) + \boldsymbol{e}_q^T(t)\boldsymbol{P}\dot{\boldsymbol{e}}_q(t) + \boldsymbol{e}_q^T(t)\gamma^{-1}\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{e}_q(t) - \gamma\boldsymbol{e}_f^T(t)\boldsymbol{e}_f(t) \\ + (\boldsymbol{e}_q^T(t)\boldsymbol{Q} + \dot{\boldsymbol{e}}_q^T(t)\delta\boldsymbol{Q})(\boldsymbol{A}_{er}\boldsymbol{e}_q(t) - \dot{\boldsymbol{e}}_q(t)) + (\boldsymbol{e}_q^T(t)\boldsymbol{A}_{er}^T - \dot{\boldsymbol{e}}_q^T(t))(\boldsymbol{Q}\boldsymbol{e}_q(t) + \delta\boldsymbol{Q}\dot{\boldsymbol{e}}_q(t)) \\ + \boldsymbol{e}_q^T(t)\boldsymbol{A}_e^T\varrho^{-1}\boldsymbol{Q}\boldsymbol{A}_e\boldsymbol{e}_q(t) + \dot{\boldsymbol{e}}_q^T(t)\delta\boldsymbol{Q}\boldsymbol{F}_r\boldsymbol{e}_f(t) + \boldsymbol{e}_f^T(t)\delta\boldsymbol{F}_r^T\boldsymbol{Q}\dot{\boldsymbol{e}}_q(t) < 0. \end{aligned} \tag{80}$$

From Eq. (80), using the notation (58), the following stability condition can be obtained

$$\dot{v}(\boldsymbol{e}_d(t)) = \boldsymbol{e}_d^T(t)\boldsymbol{P}_{de}\boldsymbol{e}_d(t) < 0, \tag{81}$$

where

$$\boldsymbol{P}_{de} = \begin{bmatrix} \boldsymbol{Q}\boldsymbol{A}_{er} + \boldsymbol{A}_{er}^T\boldsymbol{Q} + \varrho^{-1}\boldsymbol{A}_e^T\boldsymbol{Q}\boldsymbol{A}_e + \gamma^{-1}\boldsymbol{C}^T\boldsymbol{C} & \boldsymbol{P}-\boldsymbol{Q} + \delta\boldsymbol{A}_{er}^T\boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{P}-\boldsymbol{Q} + \delta\boldsymbol{Q}\boldsymbol{A}_{er} & -2\delta\boldsymbol{Q} & \delta\boldsymbol{Q}\boldsymbol{F} \\ \boldsymbol{0} & \delta\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p \end{bmatrix} < 0. \tag{82}$$

It can be easily stated using Eq. (76) that

$$\boldsymbol{Q}\boldsymbol{A}_{er} + \boldsymbol{A}_{er}^T\boldsymbol{Q} + \varrho^{-1}\boldsymbol{A}_e^T\boldsymbol{Q}\boldsymbol{A}_e = \frac{a}{\varrho}(\boldsymbol{Q}\boldsymbol{A}_e + \boldsymbol{A}_e^T\boldsymbol{Q}) + \frac{a^2-\varrho^2}{\varrho}\boldsymbol{Q} + \frac{1}{\varrho}\boldsymbol{A}_e^T\boldsymbol{Q}\boldsymbol{A}_e, \tag{83}$$

so, completing to square the elements in Eq. (83), it is immediate that

$$\boldsymbol{Q}\boldsymbol{A}_{er} + \boldsymbol{A}_{er}^T\boldsymbol{Q} + \varrho^{-1}\boldsymbol{A}_e^T\boldsymbol{Q}\boldsymbol{A}_e = (\boldsymbol{A}_e + a\boldsymbol{I}_n)^T\varrho^{-1}\boldsymbol{Q}(\boldsymbol{A}_e + a\boldsymbol{I}_n) - \varrho\boldsymbol{Q}. \tag{84}$$

Substituting Eqs. (76) and (84) in Eq. (82) gives

$$\begin{bmatrix} -\varrho\boldsymbol{Q} + (\boldsymbol{A}_e + a\boldsymbol{I}_n)^T\varrho^{-1}\boldsymbol{Q}(\boldsymbol{A}_e + a\boldsymbol{I}_n) + \gamma^{-1}\boldsymbol{C}^T\boldsymbol{C} & \boldsymbol{P}-\boldsymbol{Q} + \frac{\delta}{\varrho}\boldsymbol{A}_e^T\boldsymbol{Q} + \frac{\delta}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{P}-\boldsymbol{Q} + \frac{\delta}{\varrho}\boldsymbol{Q}\boldsymbol{A}_e + \frac{\delta}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} & -2\delta\boldsymbol{Q} & \frac{\delta}{\varrho}\boldsymbol{Q}\boldsymbol{F} \\ \boldsymbol{0} & \frac{\delta}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p \end{bmatrix} < 0 \tag{85}$$

and using twice the Schur complement property, Eq. (85) can be rewritten as

$$\begin{bmatrix} -\varrho\boldsymbol{Q} & (\boldsymbol{A}_e + a\boldsymbol{I}_n)^T\boldsymbol{Q} & \boldsymbol{P}-\boldsymbol{Q} + \frac{\delta}{\varrho}\boldsymbol{A}_e^T\boldsymbol{Q} + \frac{\delta}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} & \boldsymbol{0} & \boldsymbol{C}^T \\ \boldsymbol{Q}(\boldsymbol{A}_e + a\boldsymbol{I}_n) & -\varrho\boldsymbol{Q} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{P}-\boldsymbol{Q} + \frac{\delta}{\varrho}\boldsymbol{Q}\boldsymbol{A}_e + \frac{\delta}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} & \boldsymbol{0} & -2\delta\boldsymbol{Q} & \frac{\delta}{\varrho}\boldsymbol{Q}\boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \frac{\delta}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0. \tag{86}$$

Thus, for $\boldsymbol{A}_e$ from Eq. (30) and with the notation (59) then Eq. (86) implies Eq. (71). This concludes the proof.

Theorem 5 (*Enhanced BRL*) *The adaptive fault observer* (*19*) *and* (*20*) *is* $\mathcal{D}$*-stable if for given positive scalars* $a, \varrho \in \mathbb{R}$, $a > \varrho$, *there exist a symmetric positive definite matrix* $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, *matrices* $\boldsymbol{H} \in \mathbb{R}^{n\times p}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times m}$ *and a positive scalar* $\gamma \in \mathbb{R}$ *such that*

$$\boldsymbol{Q} = \boldsymbol{Q}^T > 0, \quad \gamma > 0, \tag{87}$$

$$\begin{bmatrix} -\varrho\boldsymbol{Q} & * & * & * \\ a\boldsymbol{Q} + \boldsymbol{Q}\boldsymbol{A} - \boldsymbol{Y}\boldsymbol{C} & -\varrho\boldsymbol{Q} & * & * \\ \boldsymbol{0} & \frac{1}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & * \\ \boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0. \tag{88}$$

$$\boldsymbol{Q}\boldsymbol{F} = \boldsymbol{C}^T\boldsymbol{H}. \tag{89}$$

When the above conditions are affirmative the estimator gain matrix can be computed by Eq. (73).

Proof. Considering that in Eq. (86) $\boldsymbol{P} = \boldsymbol{Q}$, then premultiplying the left side and postmultiplying the right side of Eq. (86) by the transformation matrix

$$\boldsymbol{T}_y = \mathrm{diag}[\,\boldsymbol{I}_n \quad \boldsymbol{I}_n \quad \delta^{-1}\boldsymbol{I}_n \quad \boldsymbol{I}_p \quad \boldsymbol{I}_m\,] \tag{90}$$

gives

$$\begin{bmatrix} -\varrho\boldsymbol{Q} & (\boldsymbol{A}_e + a\boldsymbol{I}_n)^T\boldsymbol{Q} & \frac{1}{\varrho}\boldsymbol{A}_e^T\boldsymbol{Q} + \frac{1}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} & \boldsymbol{0} & \boldsymbol{C}^T \\ \boldsymbol{Q}(\boldsymbol{A}_e + a\boldsymbol{I}_n) & -\varrho\boldsymbol{Q} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \frac{1}{\varrho}\boldsymbol{Q}\boldsymbol{A}_e + \frac{1}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} & \boldsymbol{0} & -2\delta^{-1}\boldsymbol{Q} & \frac{1}{\varrho}\boldsymbol{Q}\boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0. \tag{91}$$

Then, using the Schur complement property, the inequality (91) can be rewritten as

$$\begin{aligned} &\begin{bmatrix} -\varrho\boldsymbol{Q} & (\boldsymbol{A}_e + a\boldsymbol{I}_n)^T\boldsymbol{Q} \\ \boldsymbol{Q}(\boldsymbol{A}_e + a\boldsymbol{I}_n) & -\varrho\boldsymbol{Q} \end{bmatrix} + \begin{bmatrix} \frac{1}{\varrho}\boldsymbol{A}_e^T\boldsymbol{Q} + \frac{1}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} \\ \boldsymbol{0} \end{bmatrix}\frac{\delta}{2}\boldsymbol{Q}^{-1}\begin{bmatrix} \frac{1}{\varrho}\boldsymbol{Q}\boldsymbol{A}_e + \frac{1}{\varrho}\frac{a^2-\varrho^2}{2}\boldsymbol{Q} & \boldsymbol{0} \end{bmatrix} \\ &\quad + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \frac{1}{\varrho}\boldsymbol{Q}\boldsymbol{F} \end{bmatrix}\gamma^{-1}\boldsymbol{I}_p\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} \end{bmatrix} + \begin{bmatrix} \boldsymbol{C}^T \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}\gamma^{-1}\boldsymbol{I}_m[\,\boldsymbol{C} \quad \boldsymbol{0} \quad \boldsymbol{0} \quad \boldsymbol{0}\,] < 0. \end{aligned} \tag{92}$$

Since the second matrix element in Eq. (92) is zero matrix if $\delta = 0$ and nonzero components of the elements in the second raw are positive semidefinite, it can claim that

$$\begin{bmatrix} -\varrho\boldsymbol{Q} & (\boldsymbol{A}_e + a\boldsymbol{I}_n)^T\boldsymbol{Q} \\ \boldsymbol{Q}(\boldsymbol{A}_e + a\boldsymbol{I}_n) & -\varrho\boldsymbol{Q} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ \frac{1}{\varrho}\boldsymbol{Q}\boldsymbol{F} \end{bmatrix}\gamma^{-1}\boldsymbol{I}_p\begin{bmatrix} \boldsymbol{0} & \frac{1}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} \end{bmatrix} + \begin{bmatrix} \boldsymbol{C}^T \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}\gamma^{-1}\boldsymbol{I}_m[\,\boldsymbol{C} \quad \boldsymbol{0} \quad \boldsymbol{0}\,] < 0 \tag{93}$$

and so Eq. (93) implies the linear matrix inequality

$$\begin{bmatrix} -\varrho\boldsymbol{Q} & (\boldsymbol{A}_e + a\boldsymbol{I}_n)^T\boldsymbol{Q} & \boldsymbol{0} & \boldsymbol{C}^T \\ \boldsymbol{Q}(\boldsymbol{A}_e + a\boldsymbol{I}_n) & -\varrho\boldsymbol{Q} & \frac{1}{\varrho}\boldsymbol{Q}\boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{\varrho}\boldsymbol{F}^T\boldsymbol{Q} & -\gamma\boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0. \tag{94}$$

Thus, using Eq. (59) then Eq. (94) implies Eq. (88). This concludes the proof.

http://dx.doi.org/0.5772/67133

Theorem 6 *The adaptive fault observer (19) and (20) is $\mathcal{D}$-stable if for given positive scalars $a, \varrho \in \mathbb{R}$, $a > \varrho$, there exist a symmetric positive definite matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, matrices $\boldsymbol{H} \in \mathbb{R}^{n\times p}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times m}$ such that*

$$\boldsymbol{Q} = \boldsymbol{Q}^T > 0, \tag{95}$$

$$\begin{bmatrix} -\varrho\boldsymbol{Q} & * \\ a\boldsymbol{Q} + \boldsymbol{Q}\boldsymbol{A} - \boldsymbol{Y}\boldsymbol{C} & -\varrho\boldsymbol{Q} \end{bmatrix} < 0. \tag{96}$$

$$\boldsymbol{Q}\boldsymbol{F} = \boldsymbol{C}^T\boldsymbol{H}. \tag{97}$$

When the above conditions are affirmative the observer gain matrix can be computed by Eq. (73).

Proof. Considering only conditions implying from fault-free autonomous system (equivalent to $\boldsymbol{F} = \boldsymbol{0}$, $\boldsymbol{C} = \boldsymbol{0}$), then Eq. (88) implies directly Eq. (96). This concludes the proof.

Note, due to two integral quadratic constraints, setting the circle parameters to define $\mathcal{D}$-stabile region is relatively easy only for systems with single input and single output.

5. Extended design conditions

In order to be able to formulate the fault observer equations incorporating the symmetric, positive definite learning weight matrix $\boldsymbol{G}$, Eqs. (21), (29), and (30) can be rewritten compositely as

$$\begin{bmatrix} \dot{e}_q(t) \\ \dot{e}_f(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} - \boldsymbol{J}\boldsymbol{C} & \boldsymbol{F} \\ -\boldsymbol{G}\boldsymbol{H}^T\boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_q(t) \\ \boldsymbol{e}_f(t) \end{bmatrix}, \tag{98}$$

$$\boldsymbol{e}_y(t) = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_q(t) \\ \boldsymbol{e}_f(t) \end{bmatrix}. \tag{99}$$

Since Eq. (98) can rewritten as follows

$$\begin{bmatrix} \dot{e}_q(t) \\ \dot{e}_f(t) \end{bmatrix} = \left(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{H}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{e}_q(t) \\ \boldsymbol{e}_f(t) \end{bmatrix}, \tag{100}$$

introducing the notations

$$\tilde{\boldsymbol{e}}(t) = \begin{bmatrix} \boldsymbol{e}_q(t) \\ \boldsymbol{e}_f(t) \end{bmatrix}, \tilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \tilde{\boldsymbol{G}} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G} \end{bmatrix}, \tilde{\boldsymbol{J}} = \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{H}^T \end{bmatrix}, \tilde{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix}, \tag{101}$$

where $\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{G}} \in \mathbb{R}^{(n+p)\times(n+p)}$, $\tilde{\boldsymbol{J}} \in \mathbb{R}^{(n+p)\times m}$, $\tilde{\boldsymbol{C}} \in \mathbb{R}^{m\times(n+p)}$, $\tilde{\boldsymbol{e}}(t) \in \mathbb{R}^{n+p}$, then it follows

$$\dot{\tilde{\boldsymbol{e}}}(t) = (\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{G}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}})\tilde{\boldsymbol{e}}(t) = \tilde{\boldsymbol{A}}_e\tilde{\boldsymbol{e}}(t), \tag{102}$$

$$\boldsymbol{e}_y(t) = \tilde{\boldsymbol{C}}\tilde{\boldsymbol{e}}(t), \tag{103}$$

where

$$\tilde{\boldsymbol{A}}_e = \tilde{\boldsymbol{A}} - \tilde{\boldsymbol{G}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}}, \tag{104}$$

and $\tilde{\boldsymbol{e}}(t)$ is the generalized fault observer error.

It is necessary to note that, in general, the elements of the positive definite symmetric matrix $\boldsymbol{G}$ are unknown in advance, and have to be interactive set to adapt the observer error to the amplitude of the estimated faults. Of course, even this formulation does not mean the elimination of the matrix equality from the design conditions, because the matrix structure of $\tilde{\boldsymbol{A}}_e$ in principle leads to the bilinear matrix inequalities.

Theorem 7. *The adaptive fault observer (19) and (20) is stable if for a given symmetric, positive definite matrix* $\boldsymbol{G} \in \mathbb{R}^{p\times p}$ *there exist symmetric positive definite matrix* $\tilde{\boldsymbol{P}} \in \mathbb{R}^{(n+p)\times(n+p)}$ *and matrices* $\tilde{\boldsymbol{Z}} \in \mathbb{R}^{(n+p)\times(n+p)}$, $\tilde{\boldsymbol{Y}} \in \mathbb{R}^{(n+p)\times m}$ *such that*

$$\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{P}}^T > 0, \quad \tilde{\boldsymbol{P}}\tilde{\boldsymbol{G}} = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{Z}}, \tag{105}$$

$$\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}} + \tilde{\boldsymbol{A}}^T\tilde{\boldsymbol{P}} - \tilde{\boldsymbol{G}}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{C}} - \tilde{\boldsymbol{C}}^T\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{G}}^T < 0, \tag{106}$$

where $\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{G}} \in \mathbb{R}^{(n+p)\times(n+p)}$, $\tilde{\boldsymbol{C}} \in \mathbb{R}^{m\times(n+p)}$, $\tilde{\boldsymbol{J}} \in \mathbb{R}^{(n+p)\times m}$ *take the structures*

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \quad \tilde{\boldsymbol{G}} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G} \end{bmatrix}, \quad \tilde{\boldsymbol{C}} = [\boldsymbol{C} \quad \boldsymbol{0}], \quad \tilde{\boldsymbol{J}} = \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{H}^T \end{bmatrix}. \tag{107}$$

When the above conditions hold, the observer gain matrix is given by

$$\tilde{\boldsymbol{J}} = \tilde{\boldsymbol{Z}}^{-1}\tilde{\boldsymbol{Y}}. \tag{108}$$

Proof. Given $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{G}}$, $\tilde{\boldsymbol{C}}$ such that $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{C}})$ is observable, the Lyapunov function can be chosen as

$$v(\tilde{\boldsymbol{e}}(t)) = \tilde{\boldsymbol{e}}^T(t)\tilde{\boldsymbol{P}}\tilde{\boldsymbol{e}}(t) > 0, \tag{109}$$

where $\tilde{\boldsymbol{P}}$ is a positive definite matrix. Computing the first time derivative of Eq. (109), it yields

$$\dot{v}(\tilde{\boldsymbol{e}}(t)) = \dot{\tilde{\boldsymbol{e}}}^T(t)\tilde{\boldsymbol{P}}\tilde{\boldsymbol{e}}(t) + \tilde{\boldsymbol{e}}^T(t)\tilde{\boldsymbol{P}}\dot{\tilde{\boldsymbol{e}}}(t) < 0, \tag{110}$$

which can be restated, using Eq. (102), as

$$\dot{v}(\tilde{\boldsymbol{e}}(t)) = \tilde{\boldsymbol{e}}^T(t)(\tilde{\boldsymbol{A}}_e^T\tilde{\boldsymbol{P}} + \tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}}_e)\tilde{\boldsymbol{e}}(t) < 0. \tag{111}$$

By the Lyapunov stability theory, the asymptotic stability can be achieved if

$$\tilde{\boldsymbol{A}}_e^T\tilde{\boldsymbol{P}} + \tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}}_e < 0, \tag{112}$$

$$(\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{G}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}})^T\tilde{\boldsymbol{P}} + \tilde{\boldsymbol{P}}(\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{G}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}}) < 0, \tag{113}$$

respectively. It is evident that the matrix product $\tilde{\boldsymbol{P}}\tilde{\boldsymbol{G}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}}$ is bilinear with respect to the LMI variables $\tilde{\boldsymbol{P}}$ and $\tilde{\boldsymbol{J}}$. To facilitate the stability analysis, it can be written as

$$\tilde{\boldsymbol{P}}\tilde{\boldsymbol{G}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}} = \tilde{\boldsymbol{P}}\tilde{\boldsymbol{G}}\tilde{\boldsymbol{Z}}^{-1}\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}}=\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{-1}\tilde{\boldsymbol{G}}\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{J}}\tilde{\boldsymbol{C}}=\tilde{\boldsymbol{G}}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{C}}, \tag{114}$$

$$\tilde{\boldsymbol{G}}\tilde{\boldsymbol{Z}}^{-1} = \tilde{\boldsymbol{P}}^{-1}\tilde{\boldsymbol{G}}, \qquad \tilde{\boldsymbol{Z}}\tilde{\boldsymbol{J}} = \tilde{\boldsymbol{Y}}. \tag{115}$$

Thus, Eqs. (113) and (115) imply Eqs. (105) and (106). This concludes the proof.

Theorem 8 *The adaptive fault observer (19) and (20) is $\mathcal{D}$-stable if for a given symmetric, positive definite matrix $\boldsymbol{G}\in\mathbb{R}^{p\times p}$ and positive scalars $a, \varrho\in\mathbb{R}$, $a > \varrho$, if there exist a symmetric positive definite matrix $\tilde{\boldsymbol{Q}}\in\mathbb{R}^{(n+p)\times(n+p)}$ and matrices $\tilde{\boldsymbol{Z}}\in\mathbb{R}^{(n+p)\times(n+p)}$, $\tilde{\boldsymbol{Y}}\in\mathbb{R}^{(n+p)\times m}$ such that*

$$\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{Q}}^{T} > 0, \quad \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{G}} = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{Z}}, \quad \begin{bmatrix} -\varrho\tilde{\boldsymbol{Q}} & * \\ a\tilde{\boldsymbol{Q}} + \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{A}}-\tilde{\boldsymbol{G}}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{C}} & -\varrho\tilde{\boldsymbol{Q}} \end{bmatrix} < 0, \tag{116}$$

where $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{G}}$, $\tilde{\boldsymbol{C}}$, $\tilde{\boldsymbol{J}}$ are as in Eq. (107). When the above conditions are affirmative the observer gain matrix can be computed by Eq. (108).

Proof. Theorem 8, constructed as a generalization of the results giving stability conditions for adaptive fault observers, implies directly from Theorems 1 and 6. This concludes the proof.

6. Joint design strategy for FTC

It is assumed that the systems (16) and (17) are controllable, full state feedback control, combining with additive fault compensation from $f_e(t)$, is applied and an integral component part is added to eliminate steady tracking error. In this structure, the control law takes the form

$$\boldsymbol{u}(t) = -\overline{\boldsymbol{K}}\overline{\boldsymbol{q}}(t), \tag{117}$$

$$\overline{\boldsymbol{q}}^{T}(t) = [\,\boldsymbol{q}^{T}(t) \quad \boldsymbol{f}_e^{T}(t) \quad \boldsymbol{e}_w^{T}(t)\,], \tag{118}$$

$$\overline{\boldsymbol{K}} = [\,\boldsymbol{K}_q \quad \boldsymbol{K}_f \quad \boldsymbol{K}_w\,], \tag{119}$$

$$\boldsymbol{e}_w(t) = \int_0^t(\boldsymbol{w}(\tau)-\boldsymbol{y}(\tau))d\tau, \tag{120}$$

where $\boldsymbol{w}(t)$ is the reference output signal and $\overline{\boldsymbol{q}}(t)\in\mathbb{R}^{n+p+m}$, $\overline{\boldsymbol{K}}\in\mathbb{R}^{r\times(n+p+m)}$. Considering that in the fault-free regime

$$\boldsymbol{f}_e^{T}(t) = \boldsymbol{G}\boldsymbol{H}^{T}\boldsymbol{C}\boldsymbol{e}_y(t)\doteq\boldsymbol{0}, \tag{121}$$

and Eq. (120) follows directly

$$\dot{e}_w(t) = \boldsymbol{w}(t)-\boldsymbol{y}(t) = \boldsymbol{w}(t)-\boldsymbol{C}\boldsymbol{q}(t), \tag{122}$$

the systems (16) and (17), the fault estimation equation (21) and (121) can be expanded as

$$\begin{bmatrix} \dot{q}(t) \\ \dot{f}_e(t) \\ \dot{e}_w(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{q}(t) \\ \boldsymbol{f}_e(t) \\ \boldsymbol{e}_w(t) \end{bmatrix} + \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{u}(t) + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{I}_m \end{bmatrix} \boldsymbol{w}(t), \tag{123}$$

$$\boldsymbol{y}(t) = [\boldsymbol{C} \quad \boldsymbol{0} \quad \boldsymbol{0}] \begin{bmatrix} \boldsymbol{q}(t) \\ \boldsymbol{f}_e(t) \\ \boldsymbol{e}_w(t) \end{bmatrix}, \tag{124}$$

where $\boldsymbol{I}_m$ is the identity matrix of given dimension. Using the notations (118), (119), and

$$\overline{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \quad \overline{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \quad \overline{\boldsymbol{W}} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{I}_m \end{bmatrix}, \quad \overline{\boldsymbol{C}}^T = \begin{bmatrix} \boldsymbol{C}^T \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \tag{125}$$

$\overline{\boldsymbol{A}} \in \mathbb{R}^{(n+p+m)\times(n+p+m)}$, $\overline{\boldsymbol{B}} \in \mathbb{R}^{(n+p+m)\times r}$, $\overline{\boldsymbol{W}} \in \mathbb{R}^{(n+p+m)\times m}$ and $\overline{\boldsymbol{C}} \in \mathbb{R}^{m\times(n+p+m)}$, then

$$\dot{\overline{\boldsymbol{q}}}(t) = \overline{\boldsymbol{A}}\overline{\boldsymbol{q}}(t) + \overline{\boldsymbol{B}}\boldsymbol{u}(t) + \overline{\boldsymbol{W}}\boldsymbol{w}(t), \tag{126}$$

$$\boldsymbol{y}(t) = \overline{\boldsymbol{C}}\overline{\boldsymbol{q}}(t) \tag{127}$$

and applying the feedback control law (117) to the state space system in Eqs. (126) and (127), the expanded closed loop system becomes

$$\dot{\overline{\boldsymbol{q}}}(t) = \overline{\boldsymbol{A}}_c\overline{\boldsymbol{q}}(t) + \overline{\boldsymbol{W}}\boldsymbol{w}(t), \tag{128}$$

$$\boldsymbol{y}(t) = \overline{\boldsymbol{C}}\overline{\boldsymbol{q}}(t), \tag{129}$$

where the closed-loop system matrix of the expanded system is

$$\overline{\boldsymbol{A}}_c = \overline{\boldsymbol{A}} - \overline{\boldsymbol{B}}\overline{\boldsymbol{K}}. \tag{130}$$

In order to design the system with reference attenuations γ_2 and γ_∞, respectively, in the following is considered the transfer function matrix

$$\overline{\boldsymbol{G}}_w(s) = \overline{\boldsymbol{C}}(s\boldsymbol{I}_{n+p+m} - \overline{\boldsymbol{A}}_c)^{-1}\overline{\boldsymbol{B}}. \tag{131}$$

Proposition 1 *(H_2 control synthesis) The state feedback control (117) to the system (126) and (127) exists and $\|\overline{\boldsymbol{G}}_w(s)\|_2 < \gamma_2$ if for a given symmetric, positive definite matrix $\boldsymbol{G} \in \mathbb{R}^{p\times p}$ there exist symmetric positive definite matrices $\overline{\boldsymbol{V}} \in \mathbb{R}^{(n+p+m)\times(n+p+m)}$, $\overline{\boldsymbol{E}} \in \mathbb{R}^{m\times m}$, a matrix $\overline{\boldsymbol{Z}} \in \mathbb{R}^{r\times(n+p+m)}$ and a positive scalar $\eta \in \mathbb{R}$ such that*

$$\overline{\boldsymbol{V}} = \overline{\boldsymbol{V}}^T > 0, \quad \overline{\boldsymbol{E}} = \overline{\boldsymbol{E}}^T > 0, \quad tr(\overline{\boldsymbol{E}}) < \eta, \tag{132}$$

$$\begin{bmatrix} \overline{\boldsymbol{A}}\,\overline{\boldsymbol{V}} + \overline{\boldsymbol{V}}\,\overline{\boldsymbol{A}}^T - \overline{\boldsymbol{B}}\,\overline{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}^T\overline{\boldsymbol{B}}^T & * \\ \overline{\boldsymbol{B}}^T & -\boldsymbol{I}_r \end{bmatrix} < 0, \tag{133}$$

http://dx.doi.org/0.5772/67133

$$\begin{bmatrix} \overline{\boldsymbol{V}} & * \\ \overline{\boldsymbol{C}}\,\overline{\boldsymbol{V}} & \overline{\boldsymbol{E}} \end{bmatrix} > 0, \tag{134}$$

where

$$\overline{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \quad \overline{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \quad \overline{\boldsymbol{C}} = [\boldsymbol{C} \quad \boldsymbol{0} \quad \boldsymbol{0}], \tag{135}$$

$\overline{\boldsymbol{A}} \in \mathbb{R}^{(n+p+m)\times(n+p+m)}$, $\overline{\boldsymbol{B}} \in \mathbb{R}^{(n+p+m)\times r}$, $\overline{\boldsymbol{C}} \in \mathbb{R}^{m\times(n+p+m)}$.

When the above conditions hold, the control law gain is

$$\overline{\boldsymbol{K}} = \overline{\boldsymbol{Z}}\,\overline{\boldsymbol{V}}^{-1}. \tag{136}$$

Proof. Replacing in the inequality (6), the couple $(\boldsymbol{A}, \boldsymbol{B})$ by the pair $(\overline{\boldsymbol{A}}_c, \overline{\boldsymbol{B}})$ from Eqs. (125) and (130), consequently redefines the linear matrix inequality (6) as

$$(\overline{\boldsymbol{A}} - \overline{\boldsymbol{B}}\overline{\boldsymbol{K}})\overline{\boldsymbol{V}} + \overline{\boldsymbol{V}}(\overline{\boldsymbol{A}} - \overline{\boldsymbol{B}}\overline{\boldsymbol{K}})^T + \overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}^T < 0 \tag{137}$$

and so using the Schur complement property and the notation

$$\overline{\boldsymbol{Z}} = \overline{\boldsymbol{K}}\,\overline{\boldsymbol{V}}, \tag{138}$$

Eq. (137) implies Eq. (133).

Analogously, replacing in Eq. (7), the couple $(\boldsymbol{C}, \boldsymbol{V})$ by the pair $(\overline{\boldsymbol{C}}, \overline{\boldsymbol{V}})$, the objective of H_2 control is now to minimize the constraint $tr(\overline{\boldsymbol{C}}\,\overline{\boldsymbol{V}}\,\overline{\boldsymbol{C}}^T) < \gamma_2^2$.

Introducing the inequality

$$\overline{\boldsymbol{E}} > \overline{\boldsymbol{C}}\,\overline{\boldsymbol{V}}\,\overline{\boldsymbol{C}}^T = \overline{\boldsymbol{C}}\,\overline{\boldsymbol{V}}\,\overline{\boldsymbol{V}}^{-1}\overline{\boldsymbol{V}}\,\overline{\boldsymbol{C}}^T, \qquad \mathrm{tr}(\overline{\boldsymbol{E}}) = \eta, \tag{139}$$

with a new matrix variable $\overline{\boldsymbol{E}}$ being symmetric and positive definite, and using Schur complement property, then Eq. (139) implies directly Eq. (134). This concludes the proof.

Note, to obtain a feasible block structure of LMIs, the Schur complement property has to be used to rearrange Eq. (137) to obtain Eq. (133) while the dual Schur complement property is applied to modify Eq. (139) to obtain Eq. (134).

Proposition 2 *(H_∞ control synthesis) The state feedback control (117) to the systems (126) and (127) exists and $\|\overline{\boldsymbol{G}}(s)\|_\infty < \gamma_\infty$ if for a given symmetric, positive definite matrix $\boldsymbol{G} \in \mathbb{R}^{p\times p}$ there exist a symmetric positive definite matrix $\overline{\boldsymbol{S}} \in \mathbb{R}^{(n+p+m)\times(n+p+m)}$, a matrix $\overline{\boldsymbol{X}} \in \mathbb{R}^{r\times(n+p+m)}$ and a positive scalar $\gamma_\infty \in \mathbb{R}$ such that*

$$\overline{\boldsymbol{S}} = \overline{\boldsymbol{S}}^T > 0, \quad \gamma_\infty > 0, \tag{140}$$

$$\begin{bmatrix} \overline{\boldsymbol{A}}\overline{\boldsymbol{S}} + \overline{\boldsymbol{S}}\overline{\boldsymbol{A}}^T - \overline{\boldsymbol{B}}\overline{\boldsymbol{X}} - \overline{\boldsymbol{X}}^T\overline{\boldsymbol{B}}^T & * & * \\ \overline{\boldsymbol{B}}^T & -\gamma_\infty \boldsymbol{I}_r & * \\ \overline{\boldsymbol{C}}\overline{\boldsymbol{S}} & \boldsymbol{0} & -\gamma_\infty \boldsymbol{I}_m \end{bmatrix} < 0. \tag{141}$$

where

$$\overline{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \overline{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \overline{\boldsymbol{C}} = [\boldsymbol{C} \quad \boldsymbol{0} \quad \boldsymbol{0}], \tag{142}$$

$\overline{\boldsymbol{A}} \in \mathbb{R}^{(n+p+m)\times(n+p+m)}$, $\overline{\boldsymbol{B}} \in \mathbb{R}^{(n+p+m)\times r}$, and $\overline{\boldsymbol{C}} \in \mathbb{R}^{m\times(n+p+m)}$.

When the above conditions hold, the control law gain is

$$\overline{\boldsymbol{K}} = \overline{\boldsymbol{X}}\,\overline{\boldsymbol{S}}^{-1}. \tag{143}$$

Proof. Replacing in Eq. (9) the set of matrix parameters $(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{I}_w)$ by the foursome $(\overline{\boldsymbol{A}}_c, \overline{\boldsymbol{C}}, \overline{\boldsymbol{B}}, \boldsymbol{I}_r)$ and using the matrix variable $\overline{\boldsymbol{U}}$, then Eq. (9) gives

$$\begin{bmatrix} \overline{\boldsymbol{U}}\,\overline{\boldsymbol{A}}_c + \overline{\boldsymbol{A}}_c^T\overline{\boldsymbol{U}} & \overline{\boldsymbol{U}}\,\overline{\boldsymbol{D}} & \overline{\boldsymbol{C}}^T \\ \overline{\boldsymbol{B}}^T\overline{\boldsymbol{U}} & -\gamma_\infty \boldsymbol{I}_r & \boldsymbol{0} \\ \overline{\boldsymbol{C}} & \boldsymbol{0} & -\gamma_\infty \boldsymbol{I}_m \end{bmatrix} < 0. \tag{144}$$

Defining the transform matrix

$$\overline{\boldsymbol{T}} = \mathrm{diag}[\overline{\boldsymbol{S}} \quad \boldsymbol{I}_n \quad \boldsymbol{I}_m], \qquad \overline{\boldsymbol{S}} = \overline{\boldsymbol{U}}^{-1}, \tag{145}$$

and premultiplying the left side and postmultiplying the ride side of Eq. (144) by $\overline{\boldsymbol{T}}$, it yields

$$\begin{bmatrix} \overline{\boldsymbol{A}}_c\overline{\boldsymbol{S}} + \overline{\boldsymbol{S}}\,\overline{\boldsymbol{A}}_c^T & \overline{\boldsymbol{B}} & \overline{\boldsymbol{S}}\,\overline{\boldsymbol{C}}^T \\ \overline{\boldsymbol{B}}^T & -\gamma_\infty \boldsymbol{I}_r & \boldsymbol{0} \\ \overline{\boldsymbol{C}}\,\overline{\boldsymbol{S}} & \boldsymbol{0} & -\gamma_\infty \boldsymbol{I}_m \end{bmatrix} < 0. \tag{146}$$

Substituting Eq. (130) modifies the linear matrix inequality (146) as follows

$$\begin{bmatrix} (\overline{\boldsymbol{A}} - \overline{\boldsymbol{B}}\,\overline{\boldsymbol{K}})\overline{\boldsymbol{S}} + \overline{\boldsymbol{S}}(\overline{\boldsymbol{A}} - \overline{\boldsymbol{B}}^T\overline{\boldsymbol{K}}^T & \overline{\boldsymbol{B}} & \overline{\boldsymbol{S}}\overline{\boldsymbol{C}}^T \\ \overline{\boldsymbol{B}}^T & -\gamma_\infty \boldsymbol{I}_r & * \\ \overline{\boldsymbol{C}}\overline{\boldsymbol{S}} & \boldsymbol{0} & -\gamma_\infty \boldsymbol{I}_m \end{bmatrix} < 0 \tag{147}$$

and with the notation

$$\overline{\boldsymbol{X}} = \overline{\boldsymbol{K}}\,\overline{\boldsymbol{S}} \tag{148}$$

Eq. (147) implies Eq. (141). This concludes the proof.

It is now easy to formulate a joint approach for integrated design of FTC, where $\overline{\boldsymbol{q}}(t)$ is considered as in Eq. (118).

Theorem 9 *The state feedback control (117) to the systems (126) and (127) exists and* $\|\overline{\boldsymbol{G}}_w(s)\|_2 < \gamma_2$, $\|\overline{\boldsymbol{G}}_d(s)\|_\infty < \gamma_\infty$ *if for given symmetric, positive definite matrix* $\boldsymbol{G}\in\mathbb{R}^{p\times p}$ *and positive scalars* $a, \varrho\in IR$, a> ϱ, *there exist symmetric positive definite matrices* $\overline{\boldsymbol{V}}\in\mathbb{R}^{(n+p+m)\times(n+p+m)}$, $\tilde{\boldsymbol{Q}}\in\mathbb{R}^{(n+p)\times(n+p)}$, matrices $\overline{\boldsymbol{X}}\in\mathbb{R}^{r\times(n+p+m)}$, $\overline{\boldsymbol{E}}\in\mathbb{R}^{m\times m}$, $\tilde{\boldsymbol{Z}}\in\mathbb{R}^{(n+p)\times(n+p)}$, $\tilde{\boldsymbol{Y}}\in\mathbb{R}^{(n+p)\times m}$, *and a positive scalars* $\gamma_\infty, \eta\in\mathbb{R}$ *such that*

$$\overline{\boldsymbol{V}} = \overline{\boldsymbol{V}}^T > 0, \quad \tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{Q}}^T > 0, \quad \gamma_\infty > 0, \quad \eta > 0, \tag{149}$$

$$\begin{bmatrix} -\varrho\tilde{\boldsymbol{Q}} & * \\ a\tilde{\boldsymbol{Q}} + \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{G}}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{C}} & -\varrho\tilde{\boldsymbol{Q}} \end{bmatrix} < 0, \tag{150}$$

$$\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{G}} = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{Z}}, \tag{151}$$

$$\begin{bmatrix} \overline{\boldsymbol{A}}\,\overline{\boldsymbol{V}} + \overline{\boldsymbol{V}}\overline{\boldsymbol{A}}^T - \overline{\boldsymbol{B}}\,\overline{\boldsymbol{X}} - \overline{\boldsymbol{X}}^T\overline{\boldsymbol{B}}^T & * & * \\ \overline{\boldsymbol{B}}^T & -\gamma_\infty\boldsymbol{I}_r & * \\ \overline{\boldsymbol{C}}\,\overline{\boldsymbol{V}} & \boldsymbol{0} & -\gamma_\infty\boldsymbol{I}_m \end{bmatrix} < 0, \tag{152}$$

$$\begin{bmatrix} \overline{\boldsymbol{A}}\,\overline{\boldsymbol{V}} + \overline{\boldsymbol{V}}\,\overline{\boldsymbol{A}}^T - \overline{\boldsymbol{B}}\,\overline{\boldsymbol{X}} - \overline{\boldsymbol{X}}^T\overline{\boldsymbol{B}}^T & * \\ \overline{\boldsymbol{B}}^T & -\boldsymbol{I}_r \end{bmatrix} < 0, \tag{153}$$

$$\begin{bmatrix} \overline{\boldsymbol{V}} & * \\ \overline{\boldsymbol{C}}\,\overline{\boldsymbol{V}} & \overline{\boldsymbol{E}} \end{bmatrix} > 0, \qquad tr(\overline{\boldsymbol{E}}) < \eta. \tag{154}$$

where are $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{G}}$, $\tilde{\boldsymbol{C}}$, $\tilde{\boldsymbol{J}}$ *as in Eq. (107),* $\overline{\boldsymbol{A}}$, $\overline{\boldsymbol{B}}$, $\overline{\boldsymbol{C}}$ *as in Eq. (142), and* $\overline{\boldsymbol{K}}$ *as in Eq. (119).*

When the above conditions hold

$$\overline{\boldsymbol{K}} = \overline{\boldsymbol{X}}\,\overline{\boldsymbol{V}}^{-1}, \qquad \tilde{\boldsymbol{J}} = \overline{\boldsymbol{Z}}^{-1}\overline{\boldsymbol{Y}}. \tag{155}$$

Proof. Prescribing a unique solution of $\overline{\boldsymbol{K}}$ with respect to Eqs. (136) and (143), that is

$$\overline{\boldsymbol{V}} = \overline{\boldsymbol{S}}, \qquad \overline{\boldsymbol{X}} = \overline{\boldsymbol{Z}}, \tag{156}$$

then Eqs. (132)–(134) and (140) and (141) in the joint sense imply Eqs. (152)–(154).

The design conditions are complemented by the inequalities (150) and (151), the same as Eq. (116). This concludes the proof.

Note, the introduced H_2H_∞ control maximizes the H_2 norm over all state-feedback gains $\overline{\boldsymbol{K}}$ while the H_∞ norm constraint is optimized. The set of LMIs (152)–(154) is generally well conditioned and feasible and, since $\overline{\boldsymbol{A}}_c$ is a convergent matrix, it follows that the state of the closed-loop system converges uniformly to the desired value.

The main reason for the use of D-stability principle in the fault observer design is to adapt the fault observer dynamics to the dynamics of the fault tolerant control structure and the expected dynamics of faults. But the joint FTC design may not be linked to this principle.

7. Illustrative example

To illustrate the proposed method, a system whose dynamics is described by Eqs. (16) and (17) is considered with the matrix parameters [43]

$$\boldsymbol{A}=\begin{bmatrix} 1.380 & -0.208 & 6.715 & -5.676 \\ -0.581 & -4.290 & 0.000 & 0.675 \\ 1.067 & 4.273 & -6.654 & 5.893 \\ 0.048 & 4.273 & 1.343 & -2.104 \end{bmatrix},\ \boldsymbol{B}=\begin{bmatrix} 0.000 & 0.000 \\ 5.679 & 0.000 \\ 1.136 & -3.146 \\ 1.136 & 0.000 \end{bmatrix},\ \boldsymbol{F}=\begin{bmatrix} 1.400 \\ 1.504 \\ 2.233 \\ 0.610 \end{bmatrix},\ \boldsymbol{C}^T=\begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

To test the effectiveness and performance of the proposed estimators, the computations are carried out using the Matlab/Simulink environment and additional toolboxes, while the observer and controller design is performed by the linear matrix inequalities formulation using the functions of SeDuMi package [44]. The evaluation is performed in a standard condition, where the model to design the observer and the model for evaluation are the same and the simulations are performed according to the presented configuration of inputs and outputs.

Solving Eqs. (70)–(72), the fault observer design problem is solved as feasible where, with the prescribed stability region parameters $a = 7$, $\varrho = 5$ and the tuning parameter $\delta = 2$, the resulted matrix parameters are

$$\boldsymbol{P}=\begin{bmatrix} 11.1225 & 0.4148 & -3.7932 & -0.3068 \\ 0.4148 & 4.8026 & -2.6791 & -1.6972 \\ -3.7932 & -2.6791 & 4.2310 & -1.2725 \\ -0.3068 & -1.6972 & -1.2725 & 6.2685 \end{bmatrix},\boldsymbol{Q}=\begin{bmatrix} 6.4684 & 0.3600 & -3.1434 & -0.1831 \\ 0.3600 & 6.7121 & -4.6619 & -0.3100 \\ -3.1434 & -4.6619 & 5.6540 & -0.9782 \\ -0.1831 & -0.3100 & -0.9782 & 2.7161 \end{bmatrix},$$

$$\gamma = 27.9325, \boldsymbol{H}_4=\begin{bmatrix} 0.6166 \\ -1.2500 \end{bmatrix}, \boldsymbol{Y}=\begin{bmatrix} 18.4698 & -53.1426 \\ -1.7560 & -25.4990 \\ -6.6739 & 38.4146 \\ 0.5201 & 15.2858 \end{bmatrix}, \boldsymbol{J}_4=\begin{bmatrix} 3.6529 & -4.8333 \\ 0.7482 & 1.4554 \\ 1.6614 & 6.6685 \\ 1.1215 & 7.8699 \end{bmatrix},$$

$$\rho(\boldsymbol{A}_e)=\{-9.2971, -10.3524, -8.0807 \pm 0.5938i\}.$$

where $\rho(\boldsymbol{A}_e)$ is the observer system matrix eigenvalues spectrum. Using the same optional parameters (if necessary), there are obtained the observer gains for the design conditions introduced in Theorems 1–3 and 5–6, respectively, while

$$\boldsymbol{H}_1=\begin{bmatrix} 0.1198 \\ -0.0017 \end{bmatrix},\quad \boldsymbol{J}_1=\begin{bmatrix} 3.6529 & -4.8333 \\ 0.7482 & 1.4554 \\ 1.6614 & 6.6685 \\ 1.1215 & 7.8699 \end{bmatrix},\quad \rho(\boldsymbol{A}_e)=\left\{\begin{matrix} -3.4150, -5.2667 \\ -12.4523 \pm 20.2938i \end{matrix}\right\},$$

$$\boldsymbol{H}_2=\begin{bmatrix} 0.1809 \\ -0.5370 \end{bmatrix},\quad \boldsymbol{J}_2=\begin{bmatrix} 4.8866 & 3.6054 \\ 1.7146 & 2.5548 \\ 3.4303 & 15.4668 \\ 3.0854 & 8.3991 \end{bmatrix},\quad \rho(\boldsymbol{A}_e)=\left\{\begin{matrix} -1.0022, -7.4681 \\ -10.4435, -24.1301 \end{matrix}\right\},$$

$$\boldsymbol{H}_3 = \begin{bmatrix} 1.3248 \\ 0.3732 \end{bmatrix}, \quad \boldsymbol{J}_3 = \begin{bmatrix} 0.7403 & 0.6309 \\ 4.2089 & 9.4205 \\ 9.0063 & 15.8599 \\ -0.3253 & 0.7215 \end{bmatrix}, \quad \rho(\boldsymbol{A}_e) = \begin{Bmatrix} -3.4908 \pm 0.7441i \\ -8.6876 \pm 15.3550i \end{Bmatrix},$$

$$\boldsymbol{H}_5 = \begin{bmatrix} 0.4511 \\ -0.7967 \end{bmatrix}, \quad \boldsymbol{J}_5 = \begin{bmatrix} 3.4655 & -5.0754 \\ 0.6678 & 0.8357 \\ 1.5777 & 5.1656 \\ 1.3275 & 4.4003 \end{bmatrix}, \quad \rho(\boldsymbol{A}_e) = \begin{Bmatrix} -6.4021 \pm 1.6720i \\ -9.3518 \pm 0.4953i \end{Bmatrix},$$

$$\boldsymbol{H}_6 = \begin{bmatrix} 0.0232 \\ -0.0440 \end{bmatrix}, \quad \boldsymbol{J}_6 = \begin{bmatrix} 3.4682 & -4.8675 \\ 0.6900 & 1.0472 \\ 1.5956 & 5.4720 \\ 1.3283 & 4.5180 \end{bmatrix}, \quad \rho(\boldsymbol{A}_e) = \begin{Bmatrix} -6.4178 \pm 1.6979i \\ -9.4094 \pm 0.6999i \end{Bmatrix}.$$

Using an extended approach presented in Theorems 7 and 8, the effect of the learning weight on the dynamic performance of the adaptive fault observer is analyzed. Setting the weight $\boldsymbol{G}$ = 7.5 and using the optional factors as above, the resulted fault observer parameters are

$$\boldsymbol{H}_7 = \begin{bmatrix} 0.0530 \\ 0.1439 \end{bmatrix}, \quad \boldsymbol{J}_7 = \begin{bmatrix} -0.3895 & -1.7257 \\ 0.4619 & 2.3599 \\ 2.8130 & 4.5175 \\ -0.9425 & -0.3893 \end{bmatrix}, \quad \rho(\boldsymbol{A}_e) = \begin{Bmatrix} -2.8840 \pm 7.2479i \\ -3.3828 \pm 0.6281i \end{Bmatrix},$$

$$\boldsymbol{H}_8 = \begin{bmatrix} 0.2053 \\ 0.3222 \end{bmatrix}, \quad \boldsymbol{J}_8 = \begin{bmatrix} 2.8037 & -2.6847 \\ 0.3667 & 2.0494 \\ 1.0796 & 7.2600 \\ 0.5665 & 6.6486 \end{bmatrix}, \quad \rho(\boldsymbol{A}_e) = \begin{Bmatrix} -6.6626 \pm 1.3897i \\ -8.6430 \pm 2.3545i \end{Bmatrix}.$$

Separated simulations of fault estimation observer outputs are realized for system under the force mode control, with the control law given as

$$\boldsymbol{u}(t) = -\boldsymbol{K}_n \boldsymbol{q}(t) + \boldsymbol{W}_w \boldsymbol{w}(t). \tag{157}$$

Since separation principle holds and ($\boldsymbol{A}$, $\boldsymbol{B}$) is controllable, the eigenvalues of the closed-loop system matrix $\boldsymbol{A}_c = \boldsymbol{A} - \boldsymbol{BK}$ can be placed arbitrarily. Using the MATLAB function *place.m*, the gain matrix $\boldsymbol{K}$ is chosen that $\boldsymbol{A}_c$ has the eigenvalues {−1, −2, −3, −4}, i.e.,

$$\boldsymbol{K}_n = \begin{bmatrix} -0.1014 & -0.2357 & 0.0147 & 0.1030 \\ -1.1721 & -0.2466 & 0.1472 & -0.4907 \end{bmatrix}$$

and the signal gain matrix $\mathbf{W}_w$ is computed using the static decoupling principle as [45]

$$\boldsymbol{W} = -(\boldsymbol{C}(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{K})^{-1}\boldsymbol{B})^{-1} = \begin{bmatrix} 0.0024 & 0.1055 \\ -0.0957 & 0.0401 \end{bmatrix}. \tag{158}$$

To evaluate the validity of the proposed compensation control scheme, weighted sinusoidal fault signals are considered. Since a weighted sinusoidal fault is suitable for evaluating the tracking performance and the robustness of the control scheme because it reflects more than slow changes in the fault magnitude, the faults in simulations are generated using the scenario

$$f(t) = g(t)\sin(\omega t), \quad g(t) = \begin{cases} 0, & t \le t_{sa}, \\ \frac{1}{t_{sb}-t_{sa}}(t-t_{sa}), & t_{sa} < t_{sb}, \\ 1, & t_{sb} \le t_{ea}, \\ -\frac{1}{t_{eb}-t_{ea}}(t-t_{eb}), & t_{ea} < t_{eb}, \\ 0, & t \ge t_{eb}, \end{cases} \tag{159}$$

where it is adjusted $\omega = 1$ rad/s, $t_{sa} = 10$ s, $t_{sb} = 15$ s, $t_{ea} = 35$ s, $t_{eb} = 40$ s.

Then, with the desired system output vector, the initial system condition and the external disturbance are chosen as follows

$$\boldsymbol{w}^T(t) = [1 \quad 2], \quad \boldsymbol{q}(0) = \boldsymbol{0}, \quad \boldsymbol{D}^T = [0.610 \quad 2.233 \quad 1.504 \quad 1.400], \quad \sigma_d^2 = 0.01,$$

the faults estimates, obtained using the conditions from Theorems 1 to 6, are plotted in **Figures 1–6**. In all cases, the learning weight is set iteratively as $\boldsymbol{G}$ = 7.5. Simulations results obtained under the same simulation conditions, but realized by applying Theorems 7 and 8 with the prescribed weight $\boldsymbol{G}$ = 7.5, are given in **Figures 7** and **8**.

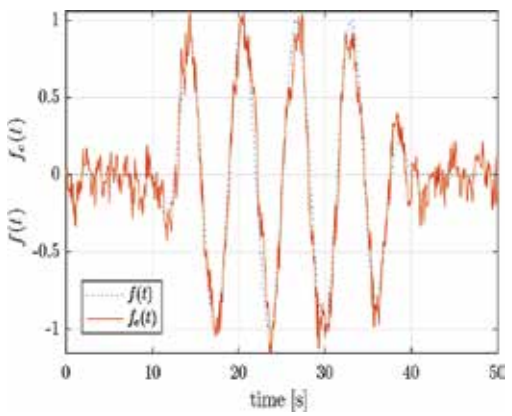


Figure 1. Estimation applying Theorem 1.

http://dx.doi.org/0.5772/67133

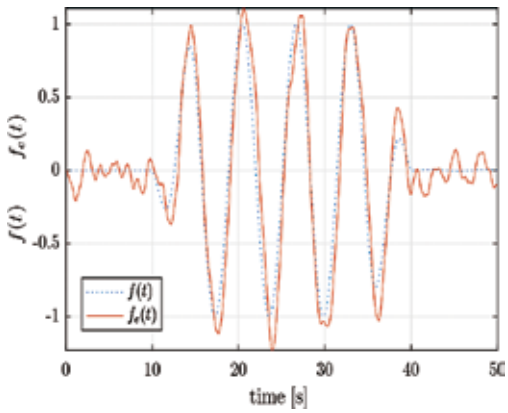


Figure 2. Estimation applying Theorem 2.

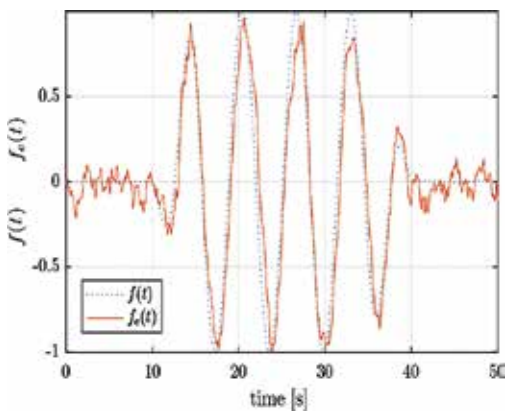


Figure 3. Estimation applying Theorem 3.

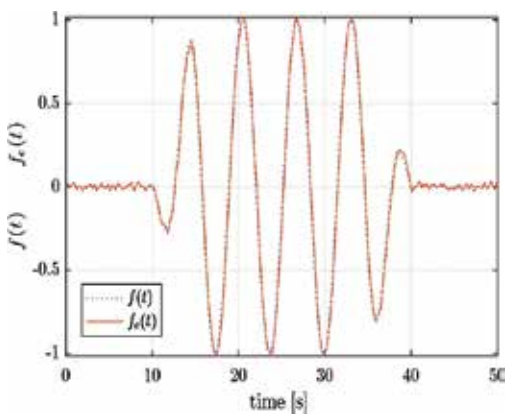


Figure 4. Estimation applying Theorem 4.

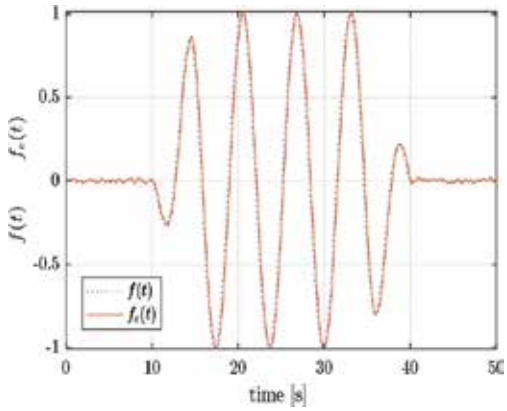


Figure 5. Estimation applying Theorem 5.

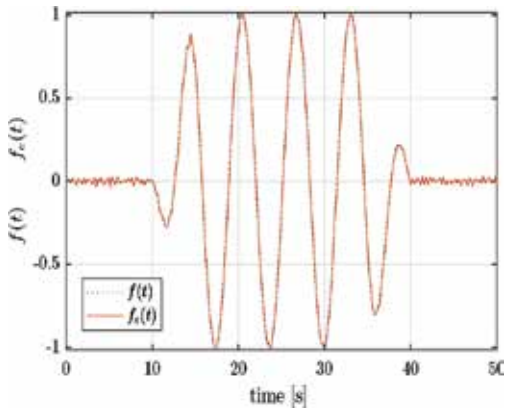


Figure 6. Estimation applying Theorem 6.

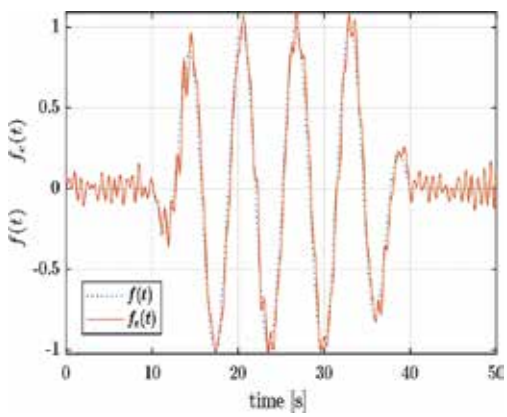


Figure 7. Estimation applying Theorem 7.

http://dx.doi.org/0.5772/67133

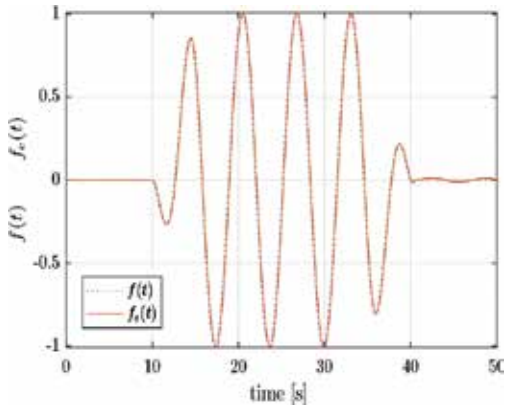


Figure 8. Estimation applying Theorem 8.

From these figures, it can be seen that fault estimation errors fast enough converge using an adaptive fault observer. Further, the extended approach with a prescribed circle $\mathcal{D}$-stability region is also effective in suppressing the disturbance noise effect on fault estimates.

Considering in the following an unforced system (126) and (127) and solving the set of linear matrix inequalities (132)–(135) to design FTC system parameters, the solution is obtained as follows

$$\overline{\boldsymbol{V}} = \begin{bmatrix} 0.1995 & 0.0196 & -0.2602 & -0.1462 & 0.0794 & 0.2031 & -0.0932 \\ 0.0196 & 1.4771 & 0.1384 & 0.2529 & -0.0064 & 0.0036 & 0.3429 \\ -0.2602 & 0.1384 & 1.4776 & 0.6864 & -0.3175 & 0.1439 & 0.5436 \\ -0.1462 & 0.2529 & 0.6864 & 0.9270 & -0.0000 & 0.0696 & 0.6344 \\ 0.0794 & -0.0064 & -0.3175 & -0.0000 & 1.4436 & 0.0080 & -0.0224 \\ 0.2031 & 0.0036 & 0.1439 & 0.0696 & 0.0080 & 2.0837 & 0.0695 \\ -0.0932 & 0.3429 & 0.5436 & 0.6344 & -0.0224 & 0.0695 & 2.1627 \end{bmatrix},$$

$$\overline{\boldsymbol{Z}} = \begin{bmatrix} -0.1746 & -0.9058 & 0.8639 & 1.1472 & 0.3790 & -0.0483 & -0.2186 \\ -1.9830 & -0.8055 & 1.9320 & -0.0805 & -1.5775 & 0.2255 & -0.3164 \end{bmatrix},$$

$$\overline{\boldsymbol{E}} = \begin{bmatrix} 3.5753 & 0.0965 \\ 0.0965 & 2.2941 \end{bmatrix}, \quad tr(\overline{\boldsymbol{E}}) = 5.8694, \quad tr(\overline{\boldsymbol{C}}\overline{\boldsymbol{V}}\overline{\boldsymbol{C}}^T) = 3.5155 < \gamma_2^2.$$

Then, the set of control law matrix parameters is

$$\boldsymbol{K}_q = \begin{bmatrix} 0.5396 & -0.8207 & 0.1959 & 1.7572 \\ -13.1540 & 0.0167 & -0.2836 & -1.9893 \end{bmatrix}, \boldsymbol{K}_f = \begin{bmatrix} 0.2652 \\ -0.4418 \end{bmatrix}, \boldsymbol{K}_w = \begin{bmatrix} -0.1308 & -0.5055 \\ 1.4821 & -0.1132 \end{bmatrix},$$

while the eigenvalue spectrum of the closed-loop system matrix is

$$\rho(\overline{\boldsymbol{A}}_c) = \{0, -0.2917, -0.4757 - 1.1533 \pm 6.7834i, -3.5221 \pm 16.1696i\}.$$

It is easy to see that the closed-loop system eigenvalues of the extended system strictly reflect the integral part of the control law that is, the set of inequalities (132)–(135) can be directly applied.

When solving the design conditions (140) and (141) for the equations of the unforced systems (126) and (127), the result is the set of matrix variables

$$\overline{\boldsymbol{S}} = \begin{bmatrix} 2.9032 & 0.2418 & -2.0149 & -0.4081 & 0.4638 & 0.2910 & -0.1012 \\ 0.2418 & 8.2299 & 1.0172 & 1.0940 & -0.0375 & 0.0847 & 1.4138 \\ -2.0149 & 1.0172 & 8.6558 & 3.9494 & -1.8552 & 1.3284 & 2.0586 \\ -0.4081 & 1.0940 & 3.9494 & 5.3411 & 0.0000 & 0.5492 & 2.6407 \\ 0.4638 & -0.0375 & -1.8552 & 0.0000 & 8.4354 & 0.3368 & -0.2675 \\ 0.2910 & 0.0847 & 1.3284 & 0.5492 & 0.3368 & 11.1517 & 0.3447 \\ -0.1012 & 1.4138 & 2.0586 & 2.6407 & -0.2675 & 0.3447 & 12.1897 \end{bmatrix},$$

$$\overline{\boldsymbol{X}} = \begin{bmatrix} -0.3425 & -4.5974 & 4.3400 & 6.8529 & 2.2149 & -0.1473 & -1.0174 \\ -13.2597 & -4.3184 & 11.1859 & -1.6190 & -9.2177 & 1.7119 & -2.1494 \end{bmatrix},$$

$$\gamma_\infty < 16.3245.$$

Based on these matrices, the closed-loop system matrix eigenvalues and the controller parameter (118) can be written out as

$$\boldsymbol{K}_q = \begin{bmatrix} 0.1556 & -0.7190 & 0.0494 & 1.5844 \\ -4.1380 & -0.3402 & 0.6571 & -1.0205 \end{bmatrix}, \boldsymbol{K}_f = \begin{bmatrix} 0.2545 \\ -0.7353 \end{bmatrix}, \boldsymbol{K}_w = \begin{bmatrix} -0.0928 & -0.3421 \\ 0.2609 & -0.0846 \end{bmatrix},$$

$$\rho(\overline{\boldsymbol{A}}_c) = \{0, -0.2054, -0.3514 - 1.2258 \pm 6.3796i, -2.1825 \pm 8.3875i\}.$$

Finally, the design conditions are designed in such a way that the upper bounds of H_2 and H_∞ norm of the system transfer function are incorporated and the parameters of the feedback controllers (117) and (118) are computed from the following set of matrix variables satisfying Eqs. (152)–(155)

$$\overline{\boldsymbol{V}} = \begin{bmatrix} 1.9774 & 0.2903 & -2.9899 & -1.0427 & 0.8321 & 0.9891 & -0.2776 \\ 0.2903 & 14.9964 & 1.0058 & 2.3203 & -0.0673 & -0.3526 & 2.3586 \\ -2.9899 & 1.0058 & 14.8053 & 5.7641 & -3.3284 & 1.0363 & 3.0673 \\ -1.0427 & 2.3203 & 5.7641 & 7.6521 & -0.0000 & 0.6069 & 4.0448 \\ 0.8321 & -0.0673 & -3.3284 & -0.0000 & 15.1342 & 0.1550 & -0.3263 \\ 0.9891 & -0.3526 & 1.0363 & 0.6069 & 0.1550 & 19.6050 & 0.4331 \\ -0.2776 & 2.3586 & 3.0673 & 4.0448 & -0.3263 & 0.4331 & 20.8419 \end{bmatrix},$$

$$\overline{\boldsymbol{X}} = \begin{bmatrix} -2.0516 & -7.0574 & 9.0166 & 12.1638 & 3.9738 & -0.1756 & -1.7169 \\ -23.6734 & -9.4744 & 21.5697 & -0.5497 & -16.5377 & 1.7345 & -3.7863 \end{bmatrix},$$

$$\overline{\boldsymbol{E}} = \begin{bmatrix} 33.4960 & 1.2672 \\ 1.2672 & 22.6309 \end{bmatrix}, \quad tr(\overline{\boldsymbol{C}}\overline{\boldsymbol{V}}\overline{\boldsymbol{C}}^T) = 30.1764 < \gamma_2^2, \quad \gamma_\infty < 22.8396,$$

while the controller matrix parameters are

$$\boldsymbol{K}_q = \begin{bmatrix} 0.0531 & -0.7236 & 0.0327 & 2.0030 \\ -14.6891 & -0.0607 & -1.2295 & -1.2123 \end{bmatrix}, \boldsymbol{K}_f = \begin{bmatrix} 0.2561 \\ -0.5650 \end{bmatrix}, \boldsymbol{K}_w = \begin{bmatrix} -0.0818 & -0.3876 \\ 0.9351 & 0.0175 \end{bmatrix}$$

and the spectrum of the closed-loop system matrix eigenvalues is

$$\rho(\overline{\boldsymbol{A}}_c) = \{0, -0.1770, -0.3009 - 1.5837 \pm 7.1369i, -5.0472 \pm 16.5305i\}.$$

Considering the same fault generation method as above, but with $\omega = 0.5$ rad/s, then for the desired system output vector, the initial system condition and the external disturbance chosen are as follows

$$\boldsymbol{w}^T(t) = [1 \quad 2], \quad \overline{\boldsymbol{q}}(0) = \boldsymbol{0}, \quad \tilde{\boldsymbol{q}}_e(0) = \boldsymbol{0}, \quad \boldsymbol{D}^T = [0.610 \quad 2.233 \quad 1.504 \quad 1.400], \quad \sigma_d^2 = 0.01,$$

the output variable responses of the closed-loop system, obtained using the conditions from Proposition 2 and Theorem 9, are shown in **Figures 9** and **10** and are stable. To the structures (141), (142), and (152)–(155), the fault estimation is designed by Eq. (116).

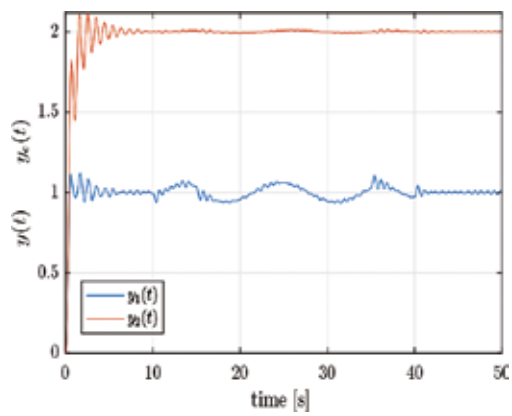


Figure 9. Compensation applying Proposition 2.

Summarizing the obtained simulation results it can be concluded that the adaptive fault estimators, designed by the standard estimation algorithm, has the worst properties (**Figure 1**) that are not significantly improved even though the conditions of synthesis are enhanced by a symmetric learning weight matrix $\boldsymbol{G}$ (**Figure 7**). Somewhat better results can be achieved when the synthesis conditions incorporate the H_∞ norm of the fault transfer function (**Figure 3**), even if they are combined with the use of an untying slack matrix $\boldsymbol{Q}$ (**Figures 2** and **5**). The best obtained results in accuracy and noise robustness are with the design conditions combining LMIs with constraints implying from $\mathcal{D}$-stability principle (**Figures 4**, **6**, and **8**).

The efficiency of the proposed algorithm to compensate the effect of an additive fault on the system output variables can be also observed. **Figures 9** and **10** show that the proposed H_2/H_∞ method increases control robustness due to the joint mixed LMI optimization that guarantees

system stability as well as the sufficient precision of compensation for a given class of slowly warring faults. Since the additive fault profile does not satisfy strictly the condition (22), its estimated time profile do not perfectly cover the actual values of the fault and where the variation of the amplitudes of $f(t)$ exceed its upper limit, there can be seen small fluctuations in compensation.

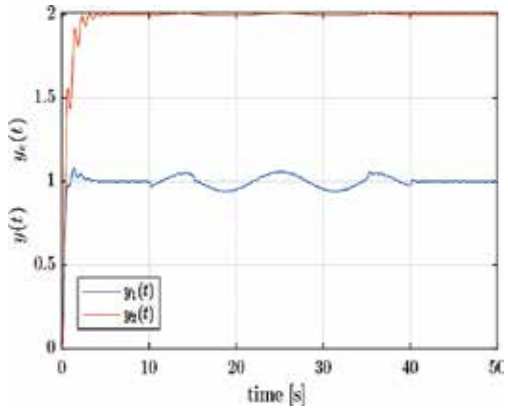


Figure 10. Compensation applying Theorem 9.

8. Concluding remarks

In this chapter, a modified approach for designing the adaptive fault observers is presented, and the $\mathcal{D}$-stability circle principle into fault observer design to outperform the two-stage known design approach in the fault observer dynamics adaptation is addressed. The design conditions are established as feasible problem, accomplishing under given quadratic constraints. Taking into consideration the slack updating effect, to cope with realistic operating conditions, the fault observer dynamics may be in the first case shifted to a stability region by exploiting the value of the tuning parameter. Integrated with the fault tolerant structures, H_2 and H_∞ norm-based analysis is carried out for compensated FTC structure to conclude about convergence of the fault compensation errors, and to derive the FTC design conditions. Using the LMI technique, the exploited mixed H_2H_∞ control design is possible to regularize the potential marginal feasibility of H_∞-norm-based conditions. Presented illustrative example confirms the effectiveness of the proposed design alternative to construct the control structure with sufficient approximation of given class slowly warring faults and compensation of their impact on the system output variables.

Acknowledgements

The work presented in this chapter was supported by VEGA, the Grant Agency of the Ministry of Education and the Academy of Science of Slovak Republic, under Grant No. 1/0348/14. This support is very gratefully acknowledged.

Author details

Dušan Krokavec*, Anna Filasová and Pavol Liščinský

*Address all correspondence to: dusan.krokavec@tuke.sk

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Košice, Slovakia

References

[1] Ding, S.X. (2013). *Model-Based Fault Diagnosis Techniques. Design Schemes, Algorithms, and Tools*, Springer-Verlag, Berlin. ISBN 978-1-4471-4798-5.

[2] Lan, J. and Patton, R.J. (2015). Integrated design of robust fault estimation and fault-tolerant control for linear systems. *Proceedings of the 54th IEEE Conference on Decision and Control CDC 2015*, Osaka, Japan, pp. 5105–5110. ISBN 978-1-4799-7886-1.

[3] Filasová, A., Gontkovič, D. and Krokavec, D. (2013). Observer-based fault estimation for linear systems with distributed time delay. *Archives of Control Sciences*, Vol. 23, No. 2, pp. 169–186. ISSN 1230-2384.

[4] Wang, H. and Yang, G.H. (2009). Integrated fault detection and control for LPV systems. *International Journal of Robust and Nonlinear Control*, Vol. 19, No. 3, pp. 341–363. ISSN 1049-8923.

[5] Tan, D. and Patton, R.J. (2015). Integrated fault estimation and fault tolerant control. A joint design. *IFAC-PapersOnLine*, Vol.48, No. 21, pp. 517–522. ISSN 1474-6670.

[6] Oucief, N., Tadjine, M. and Labiod, S. (2016). Adaptive observer-based fault estimation for a class of Lipschitz nonlinear systems. *Archives of Control Sciences*, Vol. 26, No. 2, pp. 245–259. ISSN 1230-2384.

[7] Shi, F. and Patton, R.J. (2014). Simultaneous state and fault estimation for descriptor systems using an augmented PD observer. *IFAC Proceedings Volumes*, Vol. 47, No. 3, pp. 8006–8011.

[8] Shi, F. and Patton, R.J. (2015). Fault estimation and active fault tolerant control for linear parameter varying descriptor systems. *International Journal of Robust and Nonlinear Control*, Vol. 25, No. 5, pp. 689–706. ISSN 1049-8923.

[9] Krokavec, D., Filasová, A. and Liššinský. P. (2016). Conditions with D-stability circle area in design of observer-based fault estimation. *Applied Mathematical Sciences*, Vol. 10, No. 35, pp. 1705–1717. ISSN 1312-885X.

[10] Zhang, F., Liu, G. and Fang, L. (2009). Actuator fault estimation based on adaptive H_∞ observer technique. *Proceedings of the 2009 IEEE International Conference on Mechatronics and Automation ICMA 2009*, Changchun, China, pp. 352–357. ISBN 978-1-4244-2692-8.

[11] Wang, H. and Daley, S. (1996). Actuator fault diagnosis. An adaptive observer-based technique. *IEEE Transactions on Automatic Control*, Vol. 41, No. 7, pp. 1073–1078. ISSN 0018-9286.

[12] Zhang, K. and Jiang, B. (2008). Analysis and design of adaptive fault estimation for time-varying delay systems. *Proceedings of the 27th Chinese Control Conference*, Kunming, China, pp. 38–42. ISBN 978-7-900719-70-6.

[13] Ding, X. and Frank, P. (1993). An adaptive observer-based fault detection scheme for nonlinear dynamics systems. *Proceedings of the 12th IFAC World Congres*, Sydney, Australia, pp. 307–310.

[14] Khedher, A., Benothman, K., Benrejeb, M. and Maquin, D. (2010). Adaptive observer for fault estimation in nonlinear systems described by a Takagi-Sugeno model. *Proceedings of the 18th Mediterranean Conference on Control and Automation MED'10*, Marrakech, Morocco, pp. 261–266. ISBN: 978-1-4244-8092-0.

[15] Krokavec, D. and Filasová, A. (2012). A reduced-order TS fuzzy observer scheme with application to actuator faults reconstruction. *Mathematical Problems in Engineering*, Vol. 2012, Article ID 951953, 25 p. ISSN 1024-123X.

[16] Zhang, K., Jiang, B. and Shi, P. (2013). *Observer-Based Fault Estimation and Accomodation for Dynamic Systems*, Springer-Verlag, Berlin. ISBN 978-3-642-33985-1.

[17] Francis, B.A. (1987). *A Course in H_∞ Control Theory*, Springer-Verlag, Berlin. ISBN 0-387-17069-3.

[18] Doyle, J.C., Glover, K., Khargonekar, P.P. and Francis, B.A. (1989). State-space solutions to standard H_2 and H_∞ control problems. *IEEE Transactions on Automatic Control*, Vol. 34, No. 8, pp. 831–847. ISSN 0018-9286.

[19] Rotea, M.A. and Khargonekar, P.P. (1991). H_2-optimal control with an H_∞ constraint. The state feedback case. *Automatica*, Vol. 27, No. 2, pp. 307–316. ISSN 0005-1098.

[20] Orlov, Y.V. and Aguilar, L.T. (2014). *Advanced H_∞ Control. Towards Nonsmooth Theory and Applications*, Springer Science, New York.

[21] Khargonekar, P.P. and Rotea, M.A. (1991). Mixed H_2/H_∞ control. A convex optimization approach. *IEEE Transaction on Automatic Control*, Vol. 36, No. 7, pp. 824–831. ISSN 0018-9286.

[22] Geromel, J.C., Peres, P.L.D. and Souza, S.R. (1992). Mixed H_2/H_∞ control for continuous time linear systems. *Proceedings of the 31st Conference on Decision and Control*, Tuscon, USA, pp. 3717–3722. ISBN 0-78030-872-7.

[23] Bambang, R., Shimemura, E. and Uchjida, K. (1993). Mixed H_2/H_∞ control with pole placement. State feedback case. *Proceedings of the 1993 American Control Conference*, San Francisco, USA, pp. 2777–2779. ISBN 0-78030-860-3.

[24] Scherer, C. (1995). Mixed H_2/H_∞ control. *Trends in Control. A European Perspective*, Isidori, A. (Ed.), Springer-Verlag, Berlin, pp. 173–216. TSBN 978-1-4471-3063-5.

[25] Meisami-Azad, M., Mohammadpour, J. and Grigoriadis, K.M. (2009). Upper bound mixed H_2/H_∞ control and integrated design for collocated structural systems. *Proceedings of the 2009 American Control Conference*, St. Louis, USA, pp. 4563–4568. ISBN 978-1-4244-4524-0.

[26] Goncalves, E.N., Palhares, R.M. and Takahashi, R.H.C. (2004). Multiobjective optimization applied to robust H_2/H_∞ state-feedback control synthesis. *Proceedings of the 2004 American Control Conference*, Boston, USA, pp. 4619–4624. ISBN 0-78038-335-4.

[27] Nonami, K. and Sivrioglu, S. (1996). Active vibration control using LMI-based mixed H_2/H_∞ state and output feedback control with nonlinearity. *Proceedings of the 35th Conference on Decision and Control*, Kobe, Japan, pp. 161–166. ISBN 978-0-7803-3590-5.

[28] Krokavec, D. and Filasová, A. (2015). LMI constraints on system eigenvalues placement in dynamic output control design. *Proceedings of the 2015 IEEE International Conference on Control Applications CCA 2015, Part of 2015 IEEE Multiconference on Systems and Control MSC 2015*, Sydney, Australia, pp. 1749–1754. ISBN 978-1-4799-7786-4.

[29] Zhang, K., Jiang, B. and Cocquempot, V. (2008). Adaptive observer-based fast fault estimation. *International Journal of Control, Automation and Systems*, Vol. 6, pp. 320–326. ISSN 1598-6446.

[30] Filasová, A., Krokavec, D. and Serbák, V. (2015). Application of descriptor approaches in design of PD observer-based actuator fault estimation. *Archives of Control Sciences*, Vol. 25, No. 1, pp. 51–64. ISSN 1230-2384.

[31] Fan, J., Zhang, Y. and Zheng. Z. (2011). Integrated adaptive fault diagnosis and state-feedback control for systems with constant actuator faults and control input constraints. *Proceedings ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE 2011*, Washington, USA, pp. 1–8. ISBN 978-0-7918-5480-8.

[32] Debnath, L. and Bhatta, D. (2007). *Integral Transforms and Their Applications*, Chapman & Hall, Boca Raton. ISBN 1-58488-575-0.

[33] Skelton, R.E., Iwasaki, T. and Grigoriadis, K.M. (1998). *A Unified Algebraic Approach to Control Design*, Taylor & Francis, London. ISBN 0-7484-0592-5.

[34] Doyle, J.C, Francis, B.A. and Tannenbaum, A.R. (2009). *Feedback Control Theory,* Dover Publications, Mineola. ISBN 978-0-4864-6933-1.

[35] Boyd. D. and Balakrishnan, V. (1990). A regularity result for the singular values of a transfer matrix and a qudratically convergent algorithm for computing its L_∞-norm. *Systems & Control Letters*, Vol. 15, No. 1, pp. 1–7. ISSN 0167-6911.

[36] VanAntwerp, J.G. and Braatz, R.D. (2000). A tutorial on linear and bilinear matrix inequalities. *Journal of Process Control*, Vol. 10, No. 4, pp. 363–385. ISSN 0959-1524.

[37] Haddad, W.M and Chellaboina, V. (2008). *Nonlinear Dynamical Systems and Control. A Lyapunov-Based Approach*, Princeton University Press, Princeton. ISBN 978-0-6911-3329-4.

[38] Filasová, A. and Krokavec, D. (2011). Design principles of active robust fault tolerant control systems. *Robust Control. Theory and Applications*, A. Bartoszewicz (Ed.), InTech, Rijeca, pp. 309–338. ISBN 978-953-307-229-6.

[39] Krokavec, D., Filasová, A., Liššinský. P. and Serbák, V. (2015). Design of PD observer based fault estimator using a descriptor approach. *Proceedings of the 26th International Workshop on Principles of Diagnosis DX-15*, Paris, France, pp. 235–240.

[40] Krokavec, D. and Filasová, A. (2016). On enhanced mixed H_2/H_∞ design conditions for control of linear time-invariant systems. *Proceedings of the 17th International Carpathian Control Conference ICCC 2016*, Tatranská Lomnica, Slovakia, pp. 384–389. ISBN 978-1-4673-8605-0.

[41] Patton, R.J. and Klinkhieo, S. (2009). Actuator fault estimation and compensation based on an augmented state observer approach. *Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, Shanghai, China, pp. 8482–8487. ISBN 978-1-4244-3871-6.

[42] Zhang, Y.M. and Jiang, J. (2008). Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control*, Vol. 32, pp. 229–252, ISSN 1367-5788.

[43] Kautsky, J., Nichols, N.K. and Van Dooren, P. (1985). Robust pole assignment in linear state feedback. *International Journal of Control*, Vol. 41, No. 5, pp. 1129–1155. ISSN 0020-7179.

[44] Peaucelle, D., Henrion, D., Labit, Y. and Taitz, K. (2002). *User's Guide for SeDuMi Interface 1.04*, LAAS-CNRS, Toulouse.

[45] Wang, Q.G. (2003). *Decoupling Control*, Springer-Verlag, Berlin. ISBN 3-540-44128-X.

Chapter 4

An Assessment on the Non-Invasive Methods for Condition Monitoring of Induction Motors

Muhammad Irfan, Nordin Saad, Rosdiazli Ibrahim, Vijanth S. Asirvadam, Abdullah S. Alwadie and Muhammad Aman Sheikh

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67917

Abstract

The ability to forecast motor mechanical faults at incipient stages is vital to reducing maintenance costs, operation downtime and safety hazards. This paper synthesized the progress in the research and development in condition monitoring and fault diagnosis of induction motors. The motor condition monitoring techniques are mainly classified into two categories that are invasive and non-invasive techniques. The invasive techniques are very basic, but they have some implementation difficulties and high cost. The non-invasive methods, namely MCSA, PVA and IPA, overcome the disadvantages associated to invasive methods. This book chapter reviews the various non-invasive condition monitoring methods for diagnosis of mechanical faults in induction motor and concludes that the instantaneous power analysis (IPA) and Park vector analysis (PVA) methods are best suitable for the diagnosis of small fault signatures associated to mechanical faults. Recommendations for the future research in these areas are also presented.

Keywords: condition monitoring, fault diagnosis, mechanical faults, bearing distributed faults, gear faults

1. Introduction

Induction motors are the industry workhorse due to the fact that they are rugged, reliable and economical. Induction motors are used in industry for conversion of electrical energy into mechanical energy [1]. As shown in **Figure 1**, induction motors are being used in various applications, such as in the nuclear power plants, aviation industry, transportation industry, mining industry, chemical processing plants, paper mills and the petroleum industry [1].

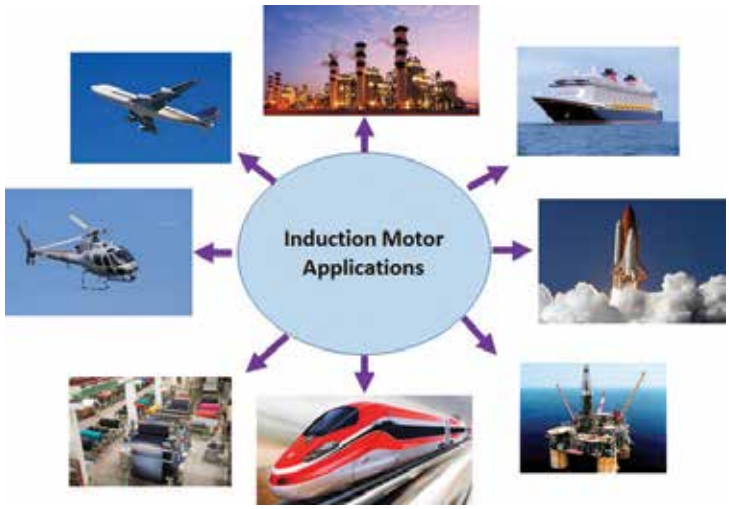


Figure 1. Applications of induction motors.

The operation of induction motor in harsh industrial environment affects its reliability and has been a critical issue. The unexpected breakdown of induction motor might result in the disturbance of critical services such as medical applications, transportation military operations and aviation. An unexpected breakdown of a motor might result in costly maintenance or loss of life in applications where continuous process is needed and where down-time is not tolerable. The induction motors require only basic maintenance and have a very low failure rate. However, the burning of the motors causes a great deal of unacceptable production loss. Consequently, diagnosing incipient faults will prevent the problems of unexpected breakdowns of the machines and it helps in reducing the maintenance costs. As reported in [2], 50% of operating cost of manufacturing and processing plants is related to maintenance. Therefore, this is a major area of concern in industries. The researchers are constantly looking for new techniques to minimize the unexpected machine failures and maintenance cost.

The scheduled replacement, scheduled maintenance and condition-based maintenance are the basic methods used in the industry for reliable operations of the machines. The scheduled replacement is a simple but expensive method as the replacement of machine parts is conducted on a regular basis. In the scheduled maintenance method, the checking and/or overhauling of the equipment has to be done on a regular basis and this method is widely used in industry as it is less expensive as compared to the scheduled replacement. The condition-based maintenance method determines the machine condition by taking measurements using sensors. By this method, the time periods between maintenance can be increased and machines can be monitored continuously so that maintenance can be scheduled on a needed basis. Condition monitoring (CM) is a method used in condition-based maintenance (CBM). It is an effective type of predictive maintenance (PM). The main steps involved in PM are shown in **Figure 2**.

Normally, the condition monitoring maintenance process would be monitoring the specific parameters like vibration, overheating, over current of equipment for early sign of coming

http://dx.doi.org/10.5772/67917

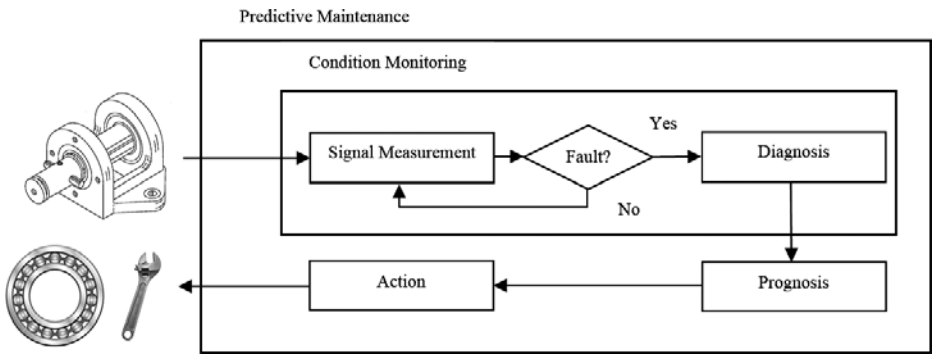


Figure 2. Activities in predictive maintenance.

failures and to predict the need of maintenance before rigorous failures. CM could be performed through visual inspection or through sophisticated fault diagnosis system. CM is suitable for continuous process plants where machine breakdowns can be very costly.

The methods of condition monitoring are categorized into two primary classifications, namely the offline tests and the online tests. The offline tests are performed by isolating the machine from main AC power supply. Although this is a comprehensive approach, sometimes causes unnecessary shutdowns on machines. Alternatively, the sensors which are installed on the machine are used to detect faults for online condition monitoring and fault diagnosis in an induction motors during the operation of the machine. The online tests cause fewer disturbances than offline tests but the results produced from online testing are more complicated and their interpretation is difficult than the offline tests. Over the past two decades, there has been an abundance of research work done in the online condition monitoring techniques for diagnosing problems in induction motors. For the detection of various faults usually affecting machines, several different techniques have been not only proposed but also used successfully. However, a good understanding of the mechanical and electrical properties of the machine in healthy and faulty conditions dramatically influences the accuracy and reliability of the online condition monitoring methods [3, 4]. Online condition monitoring techniques can be classified into two categories: firstly the classical method and secondly the digital method [4]. In classical method, electromechanical devices are used to protect the motors. The electromechanical devices are expensive, less efficient, having very slow response and not reliable as some of the devices have even shorter life than the motor itself. The digital method is the latest method for the condition monitoring and involves integrated circuits, micro-controllers, micro-processors and programmable logic controllers.

2. Overview of induction motors

An induction motor is normally composed of the following parts: frame, stator, winding, rotor and bearings. The structure of induction motor is shown in **Figure 3**. Induction motors are

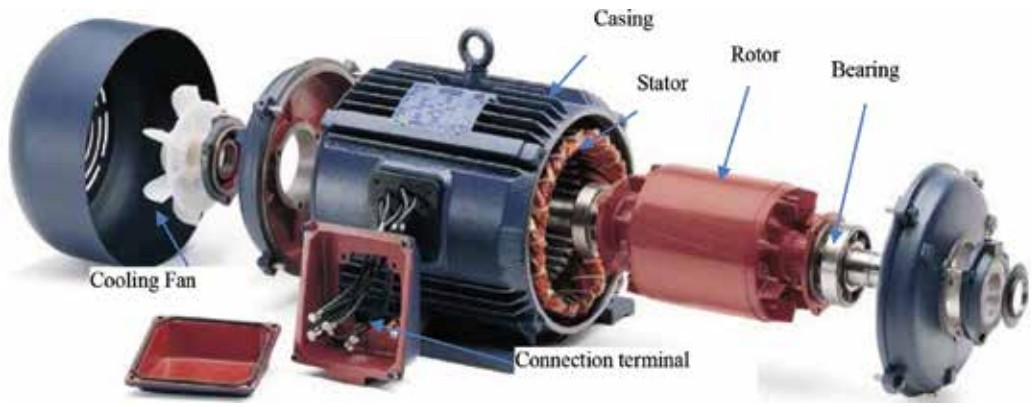


Figure 3. The structure of induction motor [5].

used in almost all types of industries. Induction motor is an asynchronous machine made up of a magnetic circuit that is joined to two electrical circuits; these circuits rotate with respect to each other. Electromagnetic induction is used to pass power from one circuit to the other. These electric motors are used to convert electrical energy into mechanical energy [5]. The conversion of energy is dependent upon the natural presence of the phenomena connecting magnetic and electrical fields in one side while motion and mechanical force are connected into the other side. On the basis of types of rotor winding, induction motors can be placed into two categories. They are the wound-rotor induction motors and squirrel-cage motors [6].

The squirrel cage induction motor is made up of conducting bars that are placed in slots of the rotor body. These conducting bars are short circuited through end rings. Magnesium, copper, or aluminium alloys are the materials usually used in manufacturing of the rotor bars. Another kind of rotor is known as a form-wound rotor since it possesses a poly-phase winding much like that of the winding of the stator. There are three slip rings which are joined to the winding of the rotor shaft. In a form-wound rotor, the slip rings are joined to a variable resistance and can restrict the current as well as the heating of the rotor [6].

The squirrel-cage induction motor is economical and robust than the wound-rotor induction motor. At constant supply of voltages and frequency, squirrel-cage induction motor runs at a constant speed. In this motor, if there is an increase in the load torque, the speed will decrease slightly. Therefore, it is appropriate to be used in drive systems that run at a constant speed [5, 6]. However, a variety of applications used in industrial areas need adjustable speeds drives. Traditionally, it is a direct current (DC) motor that is utilized in adjustable drive system. However, DC motors are expensive and possess carbon brushes that must be frequently maintained. As squirrel cage induction motors have no brushes so they are cheap and are preferred for high speed applications. Furthermore, due to the availability of solid state controllers, mostly high speed drive systems use squirrel cage induction motors. This type of induction motor is extensively utilized in drive applications of both low and high performance due to its versatility and ruggedness. Induction motors are suitable for almost all commercial and industrial applications due to their construction being so simple and they

have only a few parts, which reduce the cost of maintenance. Applications in both adjustable-speed drive (ASD) and constant-speed drive are the main uses of induction motors.

Induction motors are often exposed to operating environments that may not be ideal and in some cases are even harsh. These situations could be due to insufficient cooling, inadequate lubrication, structure vibration, overload, frequent motor starts and stops, etc. In such situations, induction motors are put under detrimental stresses which can lead to failure [7, 8]. Because of the significant role that motors play in various applications, improvement in the reliability of motors is required. The reasons why electric motors fail in industry have been commonly reported as follows [9–11]

- Wrong-rated power, voltage and current
- Mistakes during repairs
- Unstable supply voltage or current source
- Post the standard lifetime
- Overload or unbalanced load
- Electrical stress from fast switching inverters or unstable ground
- Residual stress from manufacturing
- Harsh application environment

3. Faults in induction motor

As induction motor is most often symmetrical, so faults in the motor normally disturb the symmetry of the motor. Burning of motors in the industry could be due to following reasons: thermal overloading, overloading due to undesirable stresses, air-gap eccentricity, speed oscillations, stator winding failure, broken rotor bars, bearing failure, coupled gear failures and unbalanced voltages. A concise discussion is made regarding these defects based on how important they are with regard to the condition monitoring of induction motors. According to a survey conducted in 2005 by the Electric Power Research Institute, more than 40% burning of AC motors is due to the failure of bearings (more than 50% burning of the motor is due to mechanical defects, i.e. bearing defects, gear defects, belt and pulley defects). The summary of the survey report is shown in **Figure 4** [2, 3, 12].

As stated in Peter Vas [13] and P.J. Tavner and J. Penman [14], the defects of the motor are categorized into two groups.

- Mechanical defects
- Electrical defects

Air gap eccentricities, bearing defects, shaft misalignment and abnormalities at the mechanical transmission system are included in the list of mechanical defects. Broken rotor bar and

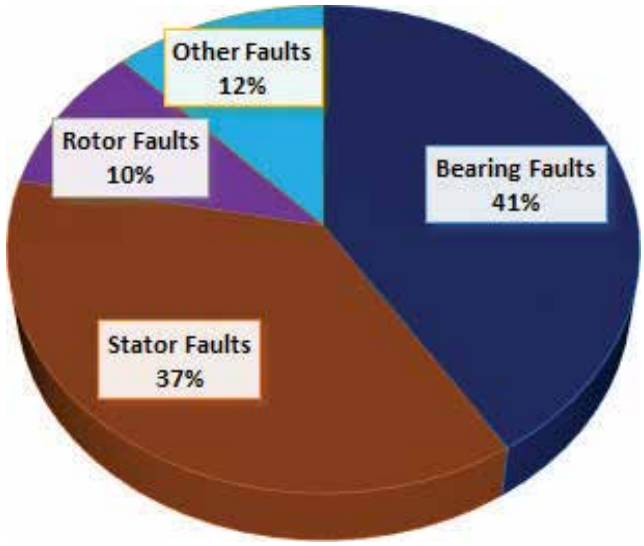


Figure 4. EPRI survey report 2005 [2, 3, 12].

stator winding defects are included in the list of electrical defects. The detail of these defects is discussed in following subsections.

3.1. Mechanical defects

The typical examples of mechanical defects that usually occur in an induction motor are presented as below.

3.1.1. Bearing defects

Bearings are widely used in rotating machinery across various industries that include paper, textile, aerospace, nuclear power plants, oil refineries, offshore pumping stations, steel, railways, construction, mining and renewable energy. The defects in the bearings cause breakdown of rotating machinery, which results in significant economic losses and in certain situations loss of human lives; for example, when a train derails or an aircraft engine fails due to a bearing defect. Bearings are typical components found in the motors that are used to allow for the shaft rotation. Majority of the motor failure is due to the bearing malfunctions. Bearings as shown in **Figure 5** are made up of inner and outer races. Several rolling elements (balls) are placed in between these two races. Cage is used to keep the balls moving at equal distance from each other. Normally, stresses developed in the motor causes fatigue in the bearing races. This fatigue causes localized defects (single-point defects like spalling or pitting or dents or holes) and distributed defects (generalized surface roughness) in the bearings of the motor [15–17].

http://dx.doi.org/10.5772/67917

Figure 5. The structure of ball bearing.

The localized defects are induced mainly due to the operational wear (contact fatigue) of bearings. If the bearings are properly installed, kept free of contamination and well lubricated during operation then bearing failure will occur only after the pre-determined life (millions of cycles) of the bearing. This type of failure initially starts in a subsurface at micro-scale level at a single-point and then due to continuous stress cycles, they eventually cause material to break. These single-point defects are characterized as spalling or pitting or localized defect [15–21]. Littman [22, 23] characterized as micro-scale subsurface defects as spalls and macro-scale surface originated defects as pitting. These types of faults produce impulsive type of vibration and serve as indication of incipient failure and due to this reason many fault diagnosis techniques have been developed to diagnose these types of faults.

The distributed defects are induced due to the manufacturing errors or due to contamination, improper lubrication, corrosion, electrical fluting and misalignment during running operation of the bearing. In these types of faults, the magnitude of the contact force between race and ball varies continuously during shaft rotation. Distributed faults create continuous type of vibration and cause premature failure of the motor bearings and thus it is important to develop a suitable condition monitoring scheme to diagnose these types of faults. **Figure 6** shows the example of typical localized and distributed defects in bearings of induction motor.

Vibration and noise are generated in the motor due to these defects [24]. The failure of bearings can also occur due to high bearing temperature. The temperature of the bearings should not increase beyond specific degrees at rated conditions. An example of this would be in the petroleum and chemical industries where the IEEE 841 standard specifies that the rise in temperature of the bearings under a rated load should not go above 45°C. Rise in the winding temperature, improper lubrication, the distribution of the temperature within the motor and

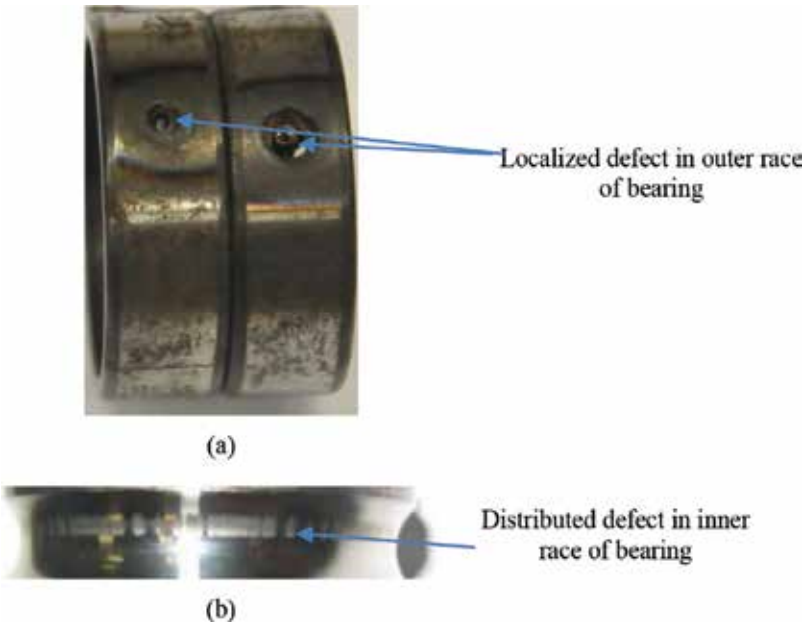


Figure 6. Example of bearing (a) localized defects and (b) distributed defects.

the operating speed of the motor are the main factors for the rise of the bearing temperature. Consequently, the bearing temperature measurements can give helpful data to estimate the health of the bearings as well as health of the motor [25, 26].

3.1.2. Air-gap eccentricity

Air-gap eccentricity is a typical defect found in the rotor of the motor. Noise and vibration in the motor structure are usually produced due to eccentricity. For the healthy motors, the centre of the rotor and stator bore is perfectly aligned. Moreover, the centre of rotation for the rotor is the same as the stator bore centre. If the rotor is not aligned centrally, radial forces or magnetic pull will be developed, which causes the rotor-stator rub resulting in the damage of rotor and stator [27]. As demonstrated in Ref. [28] air-gap eccentricities are of three kinds

- Dynamic eccentricity
- Static eccentricity
- Mixed eccentricity

A dynamic eccentricity as shown in **Figure 7** results in an unbalanced magnetic pull (UMP) that acts on the rotor. The rotation of the UMP is similar to the rotation of the motor. The UMP can be easily monitored via current or vibration analysis. On the other hand, eccentricity that is static possesses a constant pull in only one direction, which also causes a UMP. However, it is not easy to detect this type of UMP [29].

Sometimes, dynamic and static eccentricities often exist together. Moreover, one can never assume that ideal centric conditions exist. Therefore, a certain amount of eccentricity is always

http://dx.doi.org/10.5772/67917

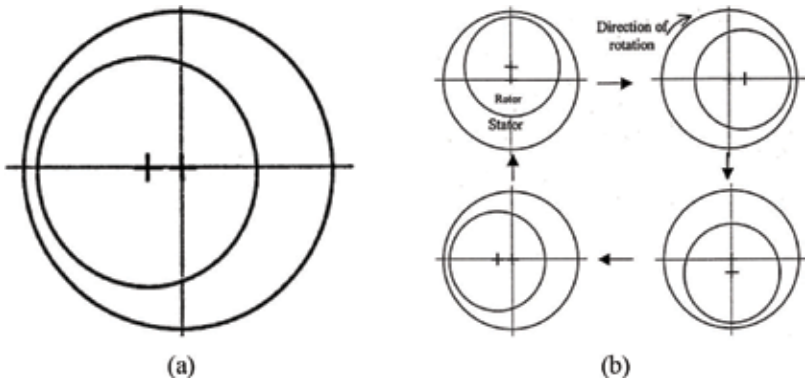


Figure 7. Example of (a) static eccentricity and (b) dynamic eccentricity [29].

expected in any real motor. This combination of eccentricities such as dynamic and static is known as mixed eccentricity.

3.1.3. Abnormalities in mechanical transmission system

Detecting abnormalities at the mechanical transmission system has been a vital area of study for quite a long time. Mechanical loads and gears are frequently connected to motors and a variety of faults such as defected gearing system, belt-pulley system and coupling misalignment are possible in these mechanical arrangements. An example of the gear fault is shown in **Figure 8**. In some applications like in aircraft, the condition monitoring of gears coupled with drive system is very much important to enhance the reliability of the gear [30]. Gear failures tend to occur when a gear is working under high stress conditions. The common gear faults are related to gear tooth irregularities namely chipped tooth, root crack, spalling, wear, pitting, tooth surface damage and broken tooth. Gear faults usually have significant effects on power transmission. They create disablement of the drives, which often causes damage to

Figure 8. Example of the gear defects [30].

other gearbox components such as the shaft, bearing, etc., by pieces of the broken tooth. The tooth breakage can be due to overload or impact and damage.

3.2. Electrical defects

The following subsections give some of the typical examples of electrical defects in induction motors.

3.2.1. Stator winding defects

The general belief is that damage in the insulation of the winding turns contributes majority of the defects related to stator. This kind of defect is known as a 'stator turn defect' [31]. In a symmetrical induction motor, a stator turn fault produces a huge amount of current to flow through turns which creates too much heat in the shorted turns. The motor will burn if this heat, which is in direct proportion to the square of the current, is more than the threshold value [32]. An example of the stator related fault is shown in **Figure 9**.

In induction motors, the insulation used in stator winding is subjected to degradation due to contamination, transient voltage stresses on the insulating material, mechanical stresses and thermal overloading. Notably, thermal stresses are the primary cause for the deterioration of the insulation in the winding of the stator. Insulation even of the best quality could experience a rapid failure if the motor is run at a temperature greater than its threshold value. Generally, the life time of the insulation is decreased by 50% for every 10°C increase over the threshold value of the temperature of stator winding [33]. Therefore, monitoring the temperature of the winding of the stator is vital so that the motor will not run at a level greater than its thermal capacity. To accomplish this, several methods have been introduced. However, these methods could not identify the exact heating-point at the earliest stage [34, 35].

Some factors that speed up the deterioration of the insulation include the defected bearings, broken rotor bars, vibrations of the rotor, movement of a coil and misalignment of the rotor and air-gap eccentricity [36]. Consequently, these mechanical failures should be identified before they cause the failure of the insulation in the stator winding [37, 38]. Another problem

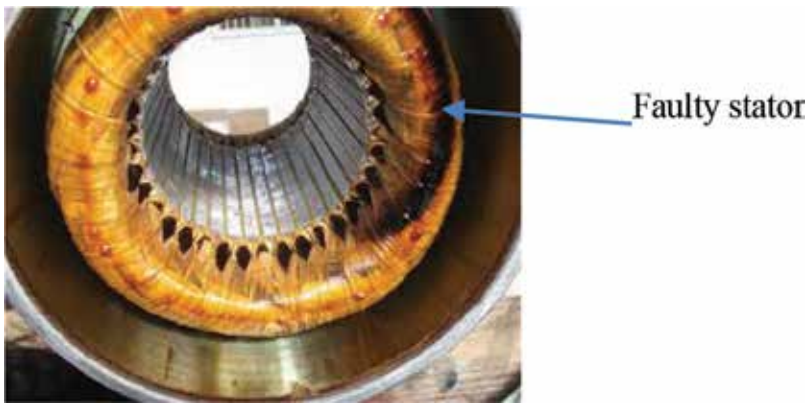


Figure 9. The stator fault in induction motor [31].

http://dx.doi.org/10.5772/67917

for the insulation is foreign materials such as dust and bearing grease which can cause contaminations over the insulation of the stator. This contamination could possibly cause the reduction in dissipation of heat from the stator winding [39]. Due to this reason it is recommended that all the motors should be kept clean and dry.

3.2.2. Rotor fault

Most often, the rotor bars in lower rated motors are produced by methods of die casting. However, the rotor bars of high rating motors are manufactured using copper. Producing rotor bars using methods of die casting has been found to cause a variety of technologically problems. Asymmetries in the rotor of the induction motors have been found which were caused by either technological problems, or because of the melting of end rings or bars. On the other hand, there is abundance of other factors causing the failure of rotors. Some of these are listed below [12, 13]

- Metallurgical stresses that are non-uniform could possibly be created in the assembly of the cage during manufacturing process and these stresses could result in a failure while the motor is operating later on.
- When thermal stresses are put on the rotor bar at the start-up of the motor, the rotor might not be capable of moving longitudinally in its slot.
- Some stresses could be developed on the rotor bars due to heavy end rings.

The reasons mentioned above could cause damage to the bars of the rotor and at the same time cause the rotor to become unbalanced. Furthermore, asymmetrical rotor currents are produced due to asymmetry on the cage of the rotor. Because of this, damage to just one rotor bar could result in damage to the surrounding bars. This damage could then spread, resulting in fractures in several rotor bars. Cracks in bars cause overheating of the bar due to which the bar may break. Consequently, the bars in the surrounding area will begin to carry higher currents subjecting them to even greater mechanical and thermal stresses. These stresses can start cracking in rotor bars and rotor lamination will be damaged as well [13]. The distribution of the temperature throughout the lamination of the rotor is also altered because of the asymmetry of the rotor. Bar fractures can take place at different areas on the rotor. During frequent starts of the motor, the chances of fracture in the rotor end rings increase [28]. Typical example of rotor faults is shown in **Figure 10**.

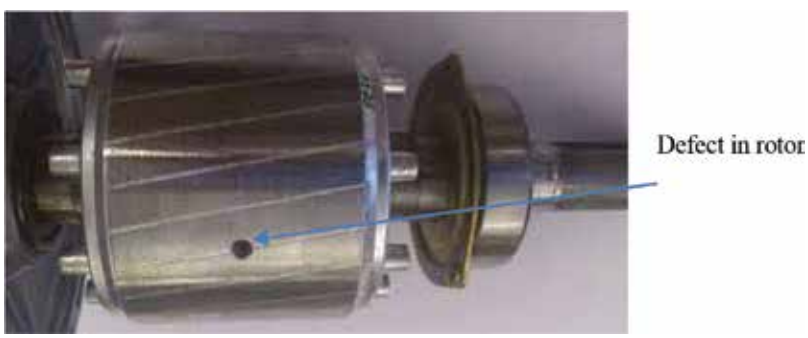


Figure 10. Example of rotor fault in induction motor [13].

4. Condition monitoring techniques

Continuously evaluating the health status of an industrial plant and its machinery throughout the entirety of its service is known as condition monitoring. Incipient failure detection is a vital process by which detection of defects in the early stages of their development is possible [13]. Fault diagnosis of the induction motor with some comprehensive condition monitoring system is becoming even more vital. An early alert about forthcoming failure is possible through the use of condition monitoring system. Furthermore, scheduling of preventive maintenance of the machines is also possible. Optimal preventive maintenance schedules are the result of this and also lead to the least amount of down-time on the machines [14]. Moreover, condition monitoring system gives indication to maintenance staff to arrange the required spare parts before serious breakdown occur on the machine, thus reduces overall down-time. Consequently, to improve productivity, reliability and safety of electric machines, a suitable condition monitoring system is essential. A tremendous significance has been put on by condition monitoring system in the environment of business because of several reasons that are listed below [13, 14]:

- To decrease the maintenance cost
- To determine the failure of machinery
- To enhance the reliability of both the machines and their parts
- To optimally use manpower and machine spare parts
- To maximize the performance of the machinery
- To enhance the failure prediction accuracy

The usage of condition monitoring for both mechanical and electrical machinery is not new. While there have been a variety of techniques developed and improved over time, invasive techniques such as acoustic emission analysis, noise analysis, thermal analysis, chemical analysis and vibration analysis, and non-invasive techniques such as motor current signature analysis (MCSA), stator current Park vector analysis and instantaneous power analysis (IPA) are considered as the most prominent methods in steady state operation conditions. The structure representing various motor faults and fault diagnosis techniques is shown in **Figure 11**.

It has been observed that even though invasive condition monitoring and fault diagnosis techniques are optimal for the diagnosis of bearing localized and distributed defects, however, costly sensors and their associated wiring is the major disadvantage of these methods [40–45]. For example, vibration sensors (accelerometers, velocity transducers) which are integral part of this technique are too much expensive. As a result, the use of invasive fault diagnosis methods is restricted in a variety of applications. This is particularly so with

http://dx.doi.org/10.5772/67917

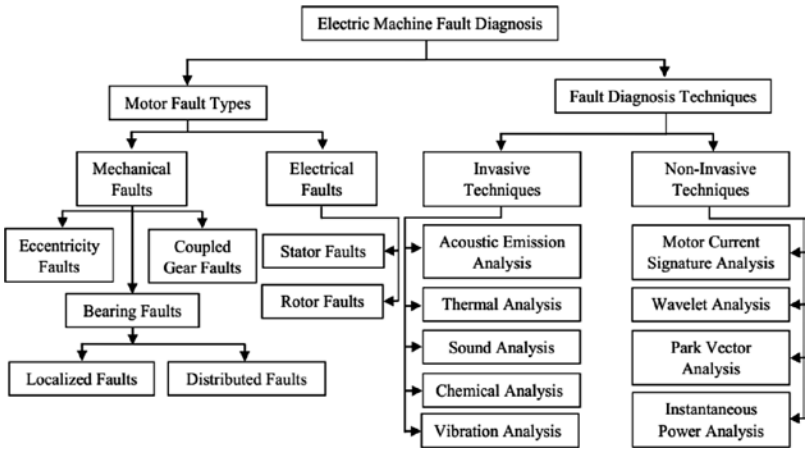


Figure 11. The structure representing various motor faults and fault diagnosis techniques.

applications using small sized motors as cost has an important role to play when making the decision as to which technique of condition monitoring is to be used. Another disadvantage of these techniques is that sensors need to be installed on the machine, and so it needs access to machine which is not possible in every application. Sensor needs to be mounted on the machine rightly for accurate results. As sensors also have some life period after which they fail, in this aspect, bearing life period is more significant than sensor life period [46–51].

Thus, the focus of this chapter is on the development of non-invasive condition monitoring and fault diagnosis method for induction motors.

5. Overview on non-invasive fault estimation parameters

Induction motor defects can be diagnosed via its terminal quantities, such as voltage, current, discharge and power, measurable outside the motor to give an indication of its condition. A fault in motor (i.e. in bearings or gears) produces a distortion of the electromagnetic field in the radial and circumferential plane due to which harmonic frequencies appear in the stator current and in instantaneous power. The stator current or instantaneous power of the healthy motor would have a single component of supply frequency (fundamental component). Motor faults will modulate the air-gap flux distribution which causes forward and backward rotating magnetic field and as a result sideband harmonic frequencies appear around fundamental component. Faults in the motor will generate one-side band below the supply frequency which is called lower-side band (LSB) and one side band above the supply frequency which is called upper-side band (USB).

Even though thermal and vibration monitoring have been utilized for decades, most of the recent researches have been directed towards electrical monitoring of the motor with emphasis on inspecting the stator current of the motor.

5.1. MCSA for bearing localized defects

The MCSA method was utilized for diagnosis of the motor and inverter defects using information from the motor stator current. In majority of applications, the stator current of an induction motor is easily obtainable because it is utilized in the protection of the motors from over-currents, ground currents. Thus, for the condition monitoring and fault diagnosis via MCSA, no extra sensors are required. It is a non-invasive method as it does not require direct access to the motor for its implementation [25, 52–58]. Thus, MCSA for the condition monitoring of the induction motors finds its application majorly in nuclear power plants, offshore pumping stations and defence industry where access to the motor is not possible.

Due to its non-invasive feature, MCSA has received the attention of many researchers and intensive research has been conducted on the MCSA. Initial efforts in MCSA can be credited to Schoen et al. [25, 53, 54]. They present a method to diagnose the motor faults based on the spectral analysis of stator current signal. Artificial neural networks were trained to learn the characteristic defect frequencies in current spectrum for the online fault diagnosis. The detection algorithm was implemented on a custom-designed test rig and its performance was verified on various fault types. Benbouzid et al. [55–57] addressed the application of MCSA for the diagnosis and localization of electrical and mechanical faults of induction motors. The initial steps taken to investigate the efficiency of MCSA for the purpose of motor fault diagnostic were discussed. Experimental results clearly illustrate the stator current spectral analysis sensitivity to induction motor faults. Later, Duque et al. [58] validated the bearing fault detection capability of MCSA on inverter-fed induction motor. In the current spectrum analysis of the motor, running values are compared with baseline values. In real time applications, baseline values are dependent on the operating conditions. To tackle this issue, Stack et al. [59] proposed new method which keeps track on baseline data at various operating conditions of the motor. They used different load conditions and for each case they compare the baseline values with running values to estimate the health condition of the motor. Along with the FFT method for the analysis of the spectrum, they used some advance signal processing and pattern recognition techniques for defecting analysis of the motor. An assessment of monitoring methods used for detection of bearing localized faults of induction motors was presented by Refs. [38, 60–62]. They proposed stator current monitoring as a very applicable technique in an industrial environment. Their research presented the current spectrum analyses along with noise cancellation method to detect mechanical faults of a variety of motors. They also propose a new method to damage the bearings via shaft current. The interpretation of the spectrum of the motor current indicates that this technique can successfully diagnose the bearing localized defects.

An improved stator current-based monitoring scheme was reported in Ref. [63], which perfectly blends Fourier transform, self-adaptive filter and rotor slot harmonics-based slip estimation techniques together. The experimental results on the bearing outer race localized

http://dx.doi.org/10.5772/67917

defects demonstrate that the presented scheme is effective. In order to diagnose the incipient faults in rotor bars, a bilinear time-frequency analysis of stator current along with adaptive filter technology was performed by Boqiang et al. [64]. The results indicate that the proposed method has the capability to diagnose broken rotor bar faults at their incipient stages. Application of MCSA for the detection of artificially damaged rolling bearings in asynchronous machine was confirmed by Trajin et al. [65]. Blodt et al. [66] and Terra et al. [67] utilized the stator current analysis and vibration analysis techniques to detect bearing localized defects. They also proposed a new model for the investigation of the effect of load torque variations on the stator current spectrum. The experimental results indicated that oscillations of the torque produce the varying frequency contents which can be observed in the stator current spectrum. Bayindir et al. [68] and Ioannides et al. [69] presented an automated fault detection system for the induction motors based on programmable logic controllers. They used the speed sensor, temperature sensor and current sensor to measure the motor speed, motor temperature and running current. The system was developed to automatic turn-off the motor if any one of the measured variable exceeds the preset values. However, the proposed condition monitoring system proves to be very expensive due to involvement of the sensors.

A novel method for the use of the stator current and efficiency of the motors as pointer of rolling-bearing defects was proposed by Frosini et al. [70, 71]. Their work describes the experimental results of bearing localized faults in outer and inner race of bearings. The analysis of the reduction in the efficiency of the motor due to defected bearings was also reported. Ebrahimi et al. [72] investigated the efficiency of the MCSA for the detection of stator faults in permanent-magnet synchronous motors. Experimental results indicate that the introduced method can detect the short-circuit fault incisively. Mehala [73] used the MCSA scheme to identify the bearing localized defects. FFT spectrum of the motor stator current was obtained and analysed. Experiments were conducted on the two defect levels (hole sizes 2 and 4 mm) in outer and inner race of the bearing under no-load and full-load conditions. The results obtained through the experiments indicate that under no-load condition, the change in amplitude values at characteristic defect frequencies is very small (<5 dB) however for full-load conditions, the change in amplitude is detectable (>5 dB to <8 dB). Along with FFT analysis of stator current, the researcher also performs wavelet and Park vector analysis of stator current for diagnosis of various motor faults.

A novel method was proposed by Romero et al. [74] which merged information entropy analysis with fuzzy logic inference to diagnose and classify faults like broken rotor bars, bearing malfunctions, stator unbalance and combinations of faults by analysing stator current signal. The proposed method shows satisfactory results that prove its suitability for online detection of single and multiple faults in an automatic way through its hardware implementation. Seera et al. [75] presented a novel approach to detect and classify comprehensive fault conditions of induction motors using a hybrid fuzzy min–max (FMM) neural network and classification and regression tree (CART). MCSA method was applied to form a database comprising stator current signatures under different motor operating conditions. Comparison of stator current analysis via FFT and wavelet transform was shown in Ref. [76]. Experimental results indicates that stator current analysis via

FFT can diagnose bearing localized defects only under steady state conditions, while, it is unable to diagnose the faults when motor is operating in transient condition. However, stator current analysis via wavelet transform has the capability to diagnose the faults under steady state and transient operating conditions of the motor. Recently, a stand-alone multi-sensor wireless system for continuous condition monitoring of induction motors has been proposed by Ref. [77]. The proposed wireless system provides a low-cost alternative to an expensive condition monitoring technology available through data acquisition equipment.

In a recent study, the fault frequencies inside the stator current spectrum were analysed through iterative method which provides an efficient fault diagnosis in the non-linear motor operations [78]. The winding function approach was used by Ref. [79] to diagnose rotor faults. The theoretical derivations were validated with experimental results. To enhance the reliability of the fault diagnosis system, Choi et al. [80] proposed a condition monitoring scheme which is based on speed feedback error management. In Ref. [81] a cross-correlation-based condition monitoring system was used to diagnose the induction motor eccentricity faults. The validity of the proposed method was confirmed through real time experiments performed on digital signal processing (DSP)-based motor drive. In a recent paper, Aydin et al. [82] proposed a hybrid approach based on multiple wireless sensor systems for fault diagnosis of induction motors. The proposed method was shown to be useful for analysing and monitoring signals from multiple induction motors. A novel analytical model based on stator current monitoring was developed by Blodt et al. [83] for the analysis of bearing defects. The results obtained through spectral analysis of the measured quantities validate the proposed theoretical approach. Stack et al. [84] presented various techniques to simulate artificial defects in the bearings of induction motor. Motor asymmetrical fault signatures were identified by Benbouzid et al. [85] through spectral analysis of motor current signal. An envelope analysis of the vibration signal was used to diagnose bearings localized defects [86]. The results achieved with proposed technique are shown to be strong function of selection signal bandwidth [86]. A wavelet-based condition monitoring system using analysis of the acoustic emission signal was proposed by Kang et al. [87]. A generalized fault analysis system in multi-phase machines was proposed by Choi et al. [88].The fault diagnosing capability of the proposed scheme has been verified mathematically through simulations. Wang et al. [89] has used an improved combination of the Hilbert and wavelet transforms to analyse the incipient bearing localized defects. Experimental results indicate that the extraction capability of bearing localized fault frequencies is greatly enhanced by the proposed method. An intelligent fault diagnosis scheme based on an independent component analysis was presented by Widodo et al. [90]. An integrated system for motor bearing fault identification is presented in Ref. [91]. Hwang et al. [91] proposed an integrated fault diagnosis system based on cepstrum coefficient method for feature extraction from motor vibration signals. The experimental results indicate the effectiveness of cepstrum in diagnosing the bearing health. Neural networks models with a decision structure are presented in Ref. [92] to analyse the bearing localized defects. The results show good performance of the implemented model and its ability to identify the bearing localized faults.

5.2. MCSA for gear defects

Gears are the main part of an electromechanical power transmission system in majority of industrial applications [93]. Unexpected failures can be prevented through detection of incipient gear faults. This will also help to reduce machine down-time and minimize financial consequences of gear damages. This cannot be realized without using an efficient fault diagnosis and condition monitoring system [94]. Although vibration analysis has been shown to be the reliable method for gear fault detection but the sensitivity to the installation position and the background noise due to external mechanical excitations are main drawbacks of the vibration measurement for gear and researchers are constantly looking to have non-invasive method for gear fault diagnosis [95, 96]. The gear fault detection using non-invasive method offers great advantages over invasive techniques principally due to its effective cost and the need of minimum changes in the system installation. In this context, an extensive research has been performed during recent years for the gear fault diagnosis using non-invasive techniques, example, using current and voltage measurements in the vicinity of motor drive systems [97–101].

As related to this, Blodt et al. [102] examined the detection of mechanical fault-related load torque oscillations in induction motors using a stator current monitoring. They developed a theoretical model to show the link between torque oscillation and modulation of a stator current component. However, they neglect the impact of the gear stiffness on the stator current. The effect of gearbox characteristic frequencies in the stator current of induction machine was studied by Ref. [95]. It was shown that the input shaft, layer shaft and output shaft frequencies in a multi-stage gearbox appear in the electromagnetic torque spectrum. Also, these harmonics appear as the sideband frequencies around the electric supply frequency of the stator current. The mesh frequencies were also well identified in the measured stator current. Their experimental investigation indicates that magnitudes of some rotation and mesh related frequencies are sensitive to the gear tooth fault [95, 103, 104]. A more rigorous attempt was carried out by Kia et al. [105, 106] to analyse the impact of motor coupled gear on the stator current spectrum. Their theoretical and experimental investigations indicate that harmonics appeared in motor current spectrum at mesh frequencies and mesh-related frequencies for healthy gears. For defected gear, additional harmonics related to fault-induced mechanical impacts are produced at the rotational frequency. The experimental results indicate that mesh and mesh-related frequencies have very weak signatures and are suppressed in the noise. The diagnosis of damaged tooth of the gear through MCSA at motor full-load conditions was reported in Refs. [107–109]. They also presented the comparison of diagnosis capabilities of vibration, acoustics and stator current analysis techniques. Stator current has been shown to be the suitable non-invasive method to diagnose gear faults. A simplified dynamic model considering a realistic behaviour of gear with minimum number of gear mechanical parameters was used to study the effect of gear torsional vibrations on motor current spectrum [110–112]. The impact of transmission error in gear was shown to be related to pinion and wheel eccentricities and tooth profile abnormalities which produces pinion and wheel rotation sidebands around the fundamental and mesh frequencies.

An attempt has been performed by Ottewill et al. [113] to diagnose the tooth pitting fault in a multi-stage gear. They have used a low-degree of freedom model for gear dynamic model. It was verified by Girsang et al. [114] that the gear teeth faults may generate mechanical impacts which can be observed in the torque and hence in machine electrical signatures. In Ref. [115], Huh et al. focused on extracting operating point independent fault signatures by using a kinematic error observer, spatial domain sampling methods and spatial domain signal filtering methods for gear fault diagnostics of electromechanical actuators. Bogiatzidis et al. [116] reported the identification of mechanical vibrations due to backlash phenomena appearing between the pinion gear and the girth gear rim of the kiln is realized using the MCSA. The proposed diagnostic method was tested on under-scale laboratory test rig. It was shown that due to fault in pinion gear, the pinion rotation frequencies appear around fundamental supply frequency.

An online gear fault diagnosis system was reported in Refs. [108, 117] using a non-invasive stator current space vector analysis. The proposed algorithm is based on the computation of the fault index by using fault-related frequencies in the current space vector instantaneous frequency spectrum. Tests performed with different load levels demonstrate a possible online condition monitoring and fault diagnosis of gear tooth surface damage fault detection. It has been demonstrated that incipient faults in gears produce small fault signatures that are sometimes buried in environment noise and thus could not be reliably detected through MCSA. Recently, a statistical tool called spectral kurtosis with reference has been proposed to define the electromechanical system healthy state reference. This technique proves its effectiveness in case of load torque oscillation fault only [118].

Although MCSA has the capability to diagnose gear defects and bearing localized defects, however, very small change in amplitude occurs at characteristic defect frequencies under no-load condition. Also in an arbitrary noise condition, the small fault signatures are buried in noise and it is difficult to discriminate between change in amplitude value due to fault and random noise [38, 53–72]. Another disadvantage of the MCSA is that sometimes two side band components (LSB and USB) lie near to the fundamental component and amplitude of the side band components is suppressed by the highest peak at fundamental component. This can create misdetection in an online fault diagnosis system [107–118].

6. IPA for motor electrical defects

Since bearings are used to support the rotor during rotation, hence any defect in bearing will affect the radial movement of rotor. Due to this radial movement, air-gap between rotor and stator will change to cause magnetic flux variations. These flux variations create modulations (oscillations) in the instantaneous power which could be observed as two sideband components, at $(2f_e - f_v)$, $(2f_e + f_v)$, and one fault component appears directly at the vibration frequency (f_v). The location of these harmonic frequencies allows the identification of abnormalities in the bearing.

The instantaneous power analysis (IPA) method has been used by several researchers to detect rotor and eccentricity defects. Ahmad [119] uses the motor current, voltage, flux and IPA methods to investigate the broken rotor bars and eccentricity defects of the motor. His

http://dx.doi.org/10.5772/67917

study shows that although the rotor bars and eccentricity defect of the motor can be detected through current, voltage and flux analysis, however, these techniques are not effective under no-load conditions of the motor. On the other hand, the instantaneous power analysis scheme diagnosis the rotor and eccentricity defects effectively even under no-load condition of the

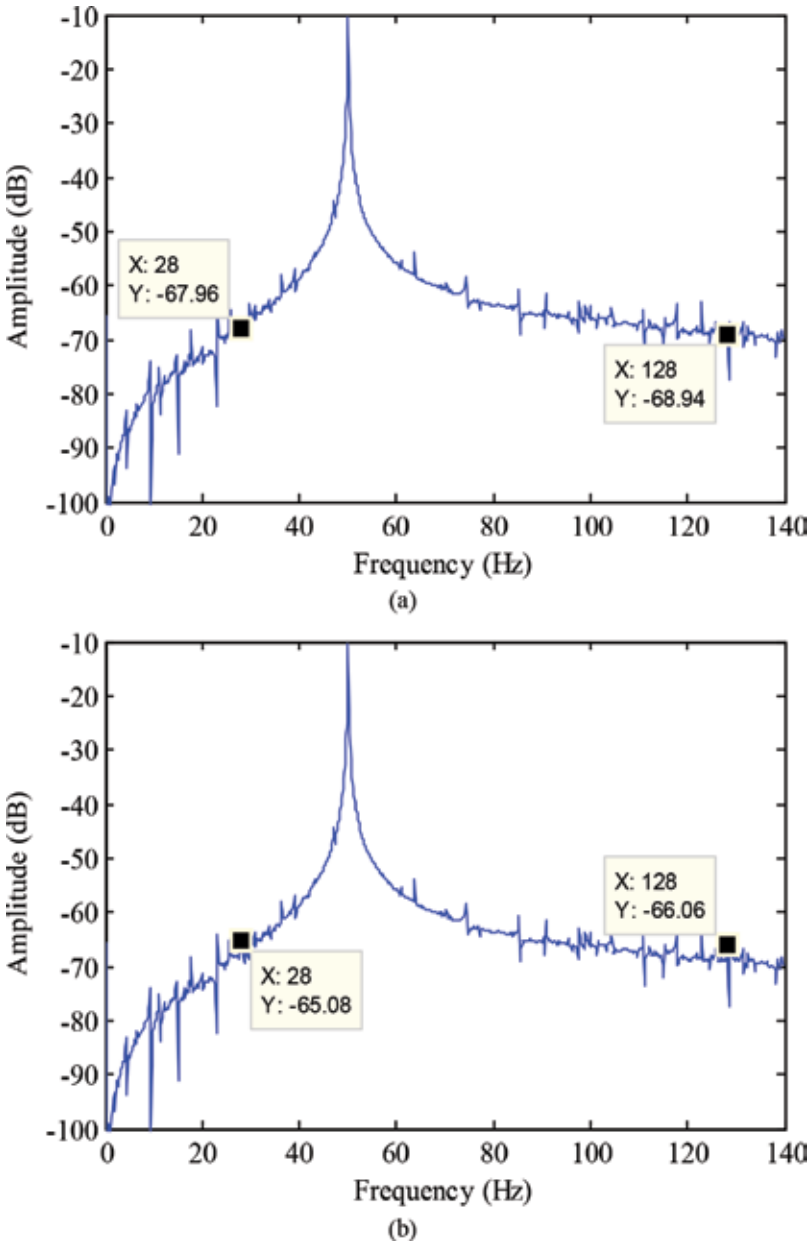


Figure 12. The stator current spectrum under no-load conditions of the motor (a) healthy motor and (b) defected bearing with 2 mm outer race defect.

motor. This is due to the reason that the motor power is the product of the supply voltage and current so it contains more information than the current and voltage only. In later work, the superiority of the IPA method over MCSA was confirmed in [120–123] for the diagnosis of faults in rotor, stator and eccentricity.

The comparison of the MCSA and IPA for 2 mm bearing outer race defect is shown in **Figures 12** and **13**, respectively. It has been observed from **Figures 12** and **13** that the MCSA is not

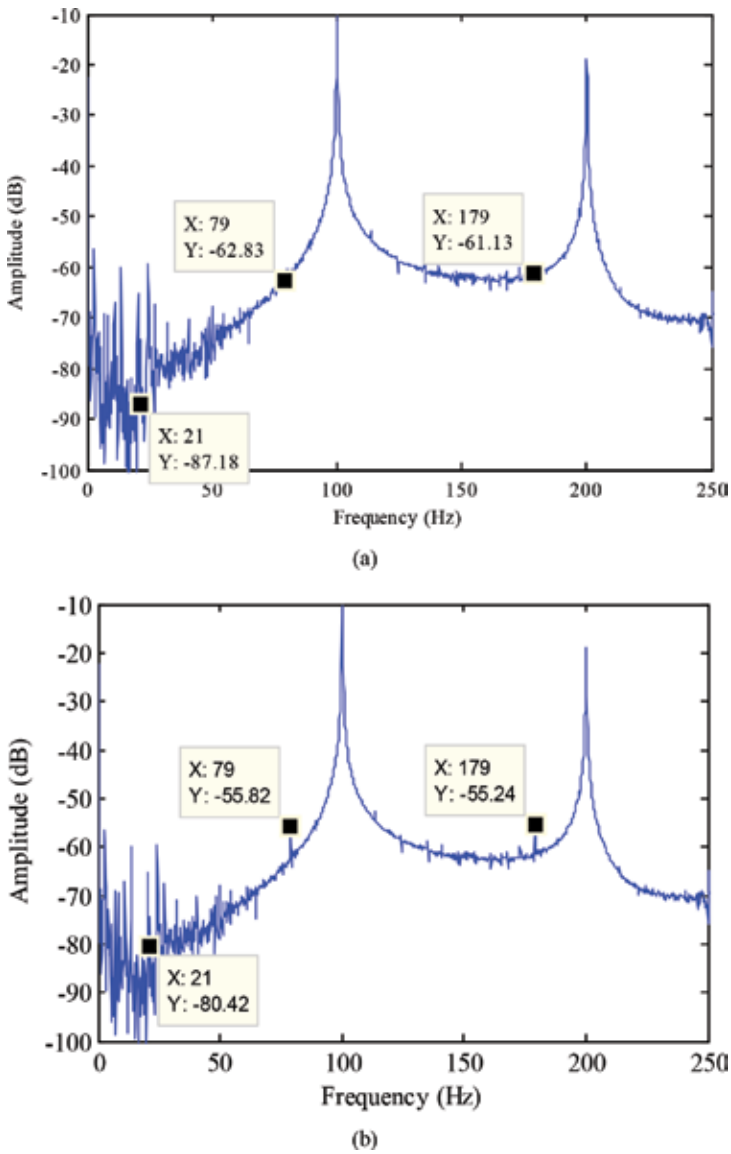


Figure 13. Instantaneous power spectrum under no-load conditions of the motor (a) healthy motor and (b) defected bearing with 2 mm outer race defect.

suitable for detection of defects at lower load levels because very small change in the amplitude value (2 dB) appears at characteristic defect frequencies of 28 and 128 Hz in stator current spectrum. However, comparatively large change in amplitude value (7 dB) appears at characteristic defect frequencies of 21, 79 and 179 Hz in instantaneous power spectrum. Also, it has been observed that the IPA carries an additional characteristic frequency component that provides an extra piece of information that can be utilized in a reliable intelligent condition monitoring system. Thus, IPA has more capability to detect bearing defects at low-load conditions as compared to MCSA.

7. Impact of environmental noise on decision making of existence of fault signatures

Incipient defects are defined as slowly developing defects or small unpredictable variations in the system. They are characterized by small amplitude compared to the useful signal. Considering the definition of incipient defects, as a slowly developing defect, it is important to address the issue of detecting these defects at the earliest possible stage meaning smallest amplitude (smallest severity). It has been observed that the low energy harmonics are produced due to incipient bearing localized faults and stator current analysis technique proves to be less efficient to detect low energy fault signatures especially under noisy environment. The IPA method has the ability to detect low energy fault signatures but in an online testing system it is difficult to discriminate the amplitude changes due to environment noise and due to existence of the fault. Therefore, it is unlikely to make decisions without considering the noise variations. As related to this, in an online fault diagnosis system the environmental noise modelling has been a practical issue and many studies have been performed to deal with it.

The averaging of multiple spectrums of stator current was used to eliminate the noise disturbances [124]. Recently, in Ref. [125], Wiener filter is used to eliminate all signatures that are irrelevant to the motor fault signatures. To achieve the high performance of the condition monitoring system, a statistical process control (SPC) is used [63]. Golafshan et al. [126, 127] presented the Singular Value Decomposition (SVD) and Hankel matrix-based de-noising process for the elimination of the background noise and the improvement in the reliability of the fault detection process. However, the proposed method is proven to be computationally intensive. A classical multiple signal classification method has been proposed by Kia et al. [128] to suppress the noise during the fault diagnosis of the motor. However, the proposed method takes long computation time to find fault signatures and is affected by the low signal-to-noise ratio. To solve this problem, an algorithm that is based on zooming in a specific frequency range of the FFT spectrum was proposed in [129]. Kim et al. [130] proposed the idea for separating rotor faults and load oscillations to reduce the false alarm rates in an online fault diagnosis system. It was reported in Ref. [131] that Wigner–Ville spectrum analysis based on cyclic spectral density (CSWVS) was able to separate the bearing fault patterns from random noise in the vibration signal. As compared with the envelope analysis for rolling element bearing diagnosis, the strongest element of CSWVS may be its diagnostic ability for bearing with

distributed faults. In the experimental study, the results of envelope analysis, no matter what filtering techniques are used, fail to extract bearing fault features and might be misinterpreted as a gear fault. However, CSWVS utilizes the second order cyclo-stationary property of the vibration signals produced by bearing distributed fault, and clearly extracts its fault features reducing the masking effect of additive stationary noise. Wang et al. [132] addressed challenging issues on de-noising and identification of the incipient fault deterioration grade from the noisy vibration signal for aircraft engine rotor prognostics. The enhanced robust methods include an adaptive wavelet de-noising technique for weak signature enhancement and correlation dimension for performance degradation assessment. The experimental results verified that the weak vibration signal features of rotor are successfully revealed and enhanced using adaptive wavelet de-noising method, and correlation well identify the rotor rub-impact fault deterioration grade.

Eccentricity faults in induction motor were analysed by Refs. [133, 134] using a cross-correlation method. The pre-determined threshold levels that have been used are as follows: amplitude less than –60 dB healthy motor, amplitude –60 dB to –40 dB tolerable defects and amplitude higher than –40 dB severe defects. A statistically derived adaptive threshold definition was proposed by Toliyat et al. [135]. The experimental results confirm the validity of the proposed approach to diagnose eccentricity and rotor fault at 70% loading condition of the motor. Although, the proposed approach considers the noise variations in real time applications; however, the performance of the proposed approach was not tested under variable operating points of the motor. In a similar work, Rajagopalan et al. [136] use an adaptive threshold scheme based on percentage of the fundamental current signal. The advanced signal processing algorithms are used to diagnose fault signatures [137]. In this study, the threshold is pre-determined based on prior tests before the diagnostic procedure starts. Although, the statistical performances of the instantaneous noise and bias with respect to motor operating points are not provided in that study but the suggested solutions present an intuitive way to design a performance oriented adaptive threshold scheme for an online fault diagnosis system. An analytical model to detect the incipient faults in the plant process in presence of an arbitrary noise was recently presented in Refs. [138–141]. Their modelled threshold scheme that is based on the probability distribution function has shown higher efficiency for incipient fault detection in noisy environment.

Notably, the pre-determined threshold schemes for reliable diagnosis of electrical faults in induction motor faults using MCSA presented in Refs. [133–137] could be used to estimate the fault severity, but these threshold schemes tend to detect sufficiently strong fault signatures to avoid noise interference. Thus the scheme could not be used to diagnose small fault signatures associated to motor mechanical faults under no-load conditions.

Fournier et al. [142] defined the threshold references in healthy condition of the motor based on the statistical spectral kurtosis measurements. The threshold scheme produces efficient results at specific speed and load levels of the motor. However, this threshold scheme does not consider instantaneous noise variations with respect to motor operating point (load, speed), so errors in the detection occur when operating point of the motor changes. Picot et al. [143] proposed a threshold scheme for the detection of small fault signatures related to bearing cage defects in permanent magnet synchronous motor (PMSM). The noise variance of

the measured motor current spectrum was minimized by integration process and threshold was designed by calculating mean and standard deviation of the frequency bins. Although, the proposed method was shown to produce good performance at various speeds (variable supply frequency) of the PMSM, however, it has some limitations. Firstly, the performance of their designed threshold was not measured at various load conditions. Secondly, the designed threshold scheme is dependent on the sampling size and window size. The large window size may attenuate the amplitude of bearing fault signatures. The impact of noise variations with respect to load variations on the reliable decision making of the existence of small fault signatures needs to be investigated. Notably, this work provides intuitive way towards the design of performance-based decision making system independent of environment impacts and motor operating points.

Limitations of the existing non-invasive fault diagnosis techniques considering environment impacts are described in **Table 1**.

Reference	Fault type	Threshold design technique	Limitations
[133, 134] 2008, 2011	Eccentricity	Pre-determined threshold	• Does not count non-linear noise variations with respect to motor speed and load • Detection with two sideband fault signatures have ambiguities
[135] 2012	Eccentricity, Rotor	Statistically derived threshold	• Does not count non-linear noise variations with respect to motor speed • Detection with two sideband fault signatures have ambiguities
[136] 2006	Eccentricity	Pre-determined threshold	• Does not count non-linear noise variations with respect to motor speed and load • Detection with two sideband fault signatures have ambiguities
[137] 2006	Bearing, Rotor	Zero input test-based statistical analysis	• Does not count non-linear noise variations with respect to motor speed and load • Detection with two sideband fault signatures have ambiguities
[142] 2013	Unbalance	Reference-based statistical analysis (spectral kurtosis)	• Does not count non-linear noise variations with respect to motor load • Detection with two sideband fault signatures have ambiguities
[143] 2014	Bearing cage Defect	Reference-based statistical analysis (mean and standard deviation)	• Does not count non-linear noise variations with respect to motor load • Detection with two sideband fault signatures have ambiguities

Table 1. Limitations of previously used MCSA-based fault diagnosis techniques considering environment impacts.

8. Park vector analysis for bearing localized defects

Park vector analysis (PVA) of motor current which is mainly used for field oriented control of induction motor has received a growing attention in the recent years in condition monitoring applications. Few researchers have used PVA with emphasis on the bearing localized faults, rotor faults, eccentricity faults and stator faults [144–147]. In Park's transformation process, the three balanced AC quantities (I_a, I_b, I_c) are reduced to two DC quantities (I_d, I_q). In analysis of three-phase synchronous machines, Park's transformation transfers three-phase stator and rotor quantities into a single rotating reference frame to eliminate the effect of time varying inductances [144, 145].

The Park vectors under ideal conditions of the motor represent a circular pattern centred at the origin. However, the shape of the circle changes in presence of the faults inside the motor and the various types of faults generate various types of shapes. Thus by detection of various patterns, one can analyse the type of fault inside the motor. The Park's transform is a graphical method to analyse the status of motor and does not require analytical expressions to calculate any specific defect frequency.

Zarie et al. [144] performed experiments on induction motor to diagnose bearing localized and extended defects (multiple holes in inner or outer race). Classification of the various faults was achieved through utilization of neural networks. It was concluded in their study that the proposed method provides a powerful and general approach to incipient fault detection. Spyropoulos et al. [145], Laughman et al. [146] and Parra et al. [148] reported that impact of the motor stator related faults could be examined through the Park vector transformation approach. It was claimed that by using this approach, effects of stator faults can be differentiated from time-varying loads. Salem et al. [147] measure the thickness of Park vector curve using curve splitting factor. The experimental results on bearing outer race defect indicate that thickness of the Park vector curve increases due to fault. Rezig et al. [149] and Salem et al. [150, 151] conducted experimental study to diagnose the bearing localized defects and eccentricity faults through Park vector transform of motor stator current. Experimental results indicate that Park vector transform method can detect the occurrence of faults but unable to identify the fault locations. In a recent study, Kuruppu et al. 152] demonstrated that Park vector analysis method has capability to diagnose stator inter-turn faults in inverter-fed field-oriented control motors. Proof of accurate fault detection capability for a wide speed range is presented through simulation and experimental results.

The example of Park vector analysis of stator current for bearing localized defects is shown in **Figure 14**. It can be observed from Figure 14 that bearing localized defects cause an increase in thickness of the I_d, I_q curves. The increase in curve thickness is due to the harmonics induced by the localized defects inside the bearing of the motor. Therefore the analysis of the I_d, I_q curves can be a useful means for the fault diagnosis of induction motors.

http://dx.doi.org/10.5772/67917

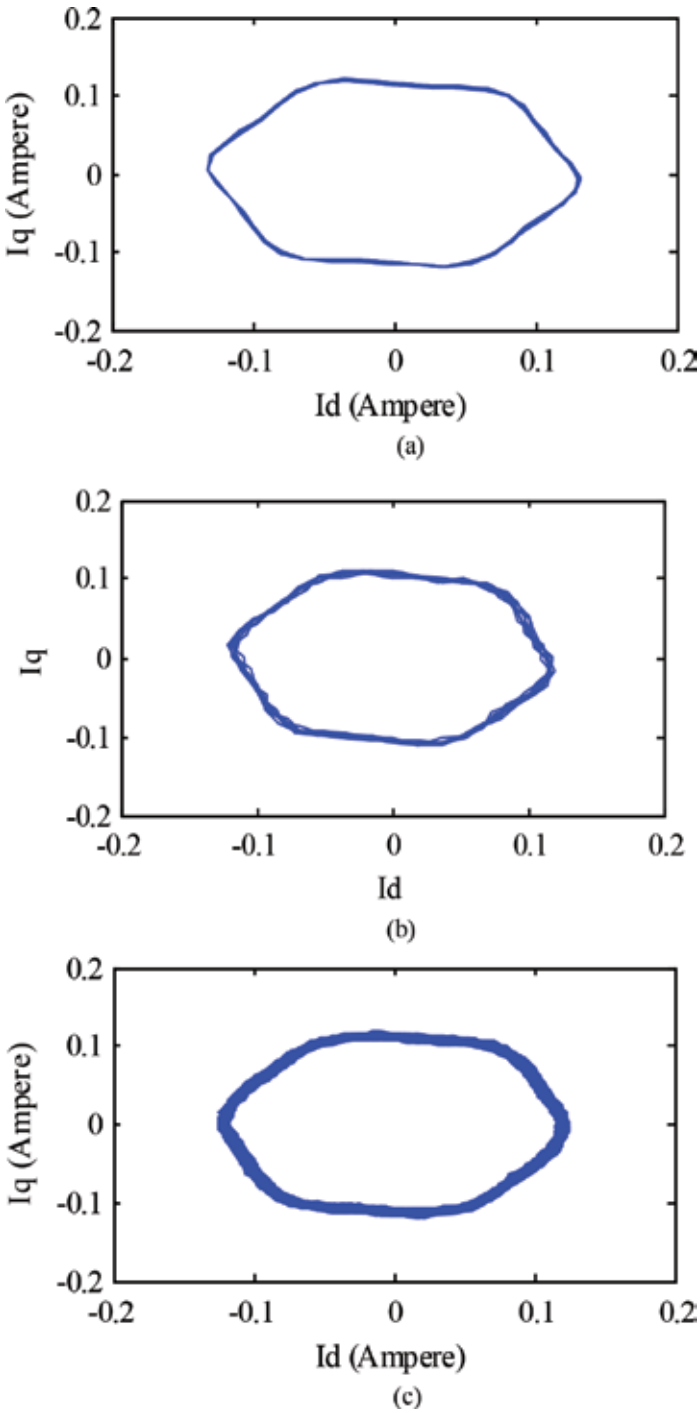


Figure 14. Example of Park vector analysis of stator current for (a) healthy bearing, (b) outer race localized defect and (c) inner race localized defect.

9. Non-invasive method for diagnosis of bearing distributed defects

Most of the published work on the bearing condition monitoring via non-invasive techniques as discussed in focused on diagnosis of bearing localized defects. It is based on some characteristic fault frequencies that appear in spectrum of the stator current or instantaneous power. However, the absence of clear characteristic fault frequencies should not be interpreted as a completely healthy condition of the bearing. On the other hand, the bearing distributed faults produce unpredictable broadband effects which are not necessarily related with specific fault frequencies. These faults are common in industry, while they are often neglected in the research literature.

Time-domain analysis is a useful feature extraction tool for condition monitoring and fault diagnosis of electrical motors. Time domain averaging (TDA) is a traditional and typical method to detect fault signals in electrical motors. It extracts a periodic component of interest from a noisy compound signal. Data-clustering techniques are used to extract an average pattern that serves as the mechanical imbalance indicator. Zhang et al. [153] proposed the idea of creating artificial distributed defects in bearings of induction motor via externally applied shaft current. This idea was important as the defects would be sorted according to the kind of signatures generated by defects instead of where the defects were located physically. Most of the condition monitoring approaches concerning fault detection found in the literature have been mainly focused on identifying single-point defects. Indeed this class of defects is quite important, however, an approach that is comprehensive and robust must have the capability of detecting not only single-point defect but also distributed defects in the bearing. The data obtained from 10 bearings failed by their proposed method, was analysed in time domain. It was shown that drastic variations in machine vibration and stator current occur in case of defected bearings. A method to segregate the bearing localized and distributed defects based on the time domain analysis of vibration, stator current and acoustic emission signals was presented by Navarro et al. [154]. The RMS values of the multi-sensory signal for healthy bearing, bearing localized and distributed defects were collected and faults were classified based on fuzzy inference analysis. In Refs. [155, 156], it was reported that generalized roughness in bearings of the motor produces a frequency spreading of the characteristic fault frequencies, thus making it difficult to detect with MCSA method. In the papers, it was proposed to use a statistical analysis of typical bearing faults in order to identify the spreading bandwidth related to bearing surface roughness faults, relying on current and vibration measurements only. A diagnostic index based on computation of energy in the previously defined bandwidth was used to diagnose bearing surface roughness faults. The proposed method was validated experimentally with vibration and current signals, with robust and reliable results. However, implementation of their proposed method needs computational efforts and implementation of this method for inner and outer race surface roughness was not given. In a recent study, Dalvand et al. [157] analyse bearing distributed defects in inner and outer race of bearing simultaneously. They use statistical analysis of the instantaneous frequency of motor voltage space vector (MVSV) and vibration signal. The statistical indices like mean value (M), RMS value, standard deviation (SD), global kurtosis (GK), skew factor (SF) and crest factor (CF) were utilized to analyse the measured signal. It was shown that mean value and RMS are not proper indices to analyse the distributed

defects in bearings because the values of these indices for healthy and defected bearings are not significantly different. However, the GK, SF and CF are proper indices to diagnose bearing distributed faults. Experimental results indicate that although the statistical time domain analysis of MVSV and vibration signals have capability to diagnose the bearing distributed defects; however the reliability of MVSV is better than vibration monitoring. The analysis of individual defect in bearing elements was not presented in their study.

10. Condition monitoring issues

The literature review has shown that the invasive methods like vibration analysis, acoustic emission, noise analysis, chemical analysis and temperature measurements were applied to protect motors from various faults. However, sensors used in these methods are very expensive. Another disadvantage of these methods is that they require access to machine for sensor installation which is not possible in every application. Also it requires special expertise for the proper installation of sensors on the machine for accurate results. Therefore sensor-based condition monitoring methods are not reliable.

Review papers presented in Refs. [158–163] on the condition monitoring techniques for the induction motors demonstrate that extensive research has been carried out in the last decade on the development of a non-invasive condition monitoring system for diagnosis of gear defects and bearing localized faults. It has been found that the mechanical vibration produce characteristic frequencies in stator current spectrum. The magnitude of the specific characteristic frequencies increases with the increase of vibration. Every type of the motor defect has its own unique characteristic defect frequencies. The MCSA has been extensively used to find out these frequency modulations.

The implementation of MCSA for motor condition monitoring is not so complex. However, low amplitude fault signatures are induced due to incipient faults under no-load condition and thus could not be reliably diagnosed through MCSA. Recent developments in non-invasive condition monitoring scheme demonstrate that PVA of motor stator current has the capability to diagnose the motor faults at incipient stages. It was shown that the proposed method would give good analysis of machine faults even if accurate fault frequency information is unavailable. In recent years, IPA method to diagnose rotor, stator and eccentricity faults at incipient stages conditions was proposed by few researchers. However, applicability of the IPA for mechanical fault diagnosis in induction motor has not been reported in literature [164–170].

The incipient defect diagnosis in the machinery with unknown distribution of measured signals and unknown changed parameters is an important issue. However, most of the fault diagnosis techniques need a priori knowledge on the signal distribution, changed parameters and the changed amplitude. Fault detection plays a key role in enhancing today's technological system's high demands for performance, productivity and security. The sensitivity of the condition monitoring and fault diagnosis methods depends on the application of main goals. When productivity is the main goal then the sensitivity requirement of condition monitoring and fault diagnosis system is weak and only large defects should be detected. However, when

security is the main goal, undetected defects even with very small severity assessment may result in catastrophic growing failures. For online fault diagnosis, when safety is the main priority, it is crucial to be able to diagnose incipient faults in presence of environment noise. Although IPA has some implementation advantages over MCSA for incipient fault detection but the signatures of IPA are affected by industrial noise perturbations. There is much scope for research to improve the capabilities of IPA in reliable online fault diagnosis in an arbitrary environment noise.

Notably, the substantial number of research studies has been found focusing into detection of bearing localized faults via non-invasive condition monitoring methods. Nonetheless, bearing distributed defects does not produce localized defect frequencies and thus could not be analysed via these methods. Recently, few studies have been found focusing on time domain analysis of stator current signal and extracting the features like RMS and kurtosis values. However, complexity involved in time domain analysis of stator current signal is the major drawback. As PVA performs the graphical analysis of stator current and does not require frequency information for fault detection thus its capability to diagnose bearing distributed defects should be further researched.

The advantages and drawbacks of the existing non-invasive condition monitoring methods are presented in **Table 2**.

No	CM technique	Advantages	Drawbacks	Reference
1.	MCSA	• Inexpensive • Non-invasive	• Unable to detect defects at incipient stages • Fault detection affected by environment noise	[25, 38, 52–73]
2.	PVA	• Inexpensive • Non-invasive	• Further investigation required to test its capability for bearing distributed defects	[144–152, 171]
3.	IPA	• Inexpensive • Non-invasive • Can detect rotor and eccentricity defects at incipient stages	• Fault detection affected by environment noise • Capability to diagnose mechanical faults needs to be tested	[119–123]

Table 2. Advantages and drawbacks of non-invasive condition monitoring methods.

11. Future directions

Based on the findings, there are further problems to be considered both in the development of the technique and in the experimental design. These include

- Development of non-invasive condition monitoring scheme to diagnose faults in an arbitrary environment noise conditions

The non-invasive IPA has shown some implementation advantages over MCSA to reliably diagnose mechanical faults. However, small fault signatures under low load conditions are buried in noise and it is difficult to discriminate amplitude changes due to fault and due to noise. Thus, there is a need to develop a robust non-invasive condition monitoring scheme to make reliable decisions on the existence of fault signatures.

- Development of non-invasive condition monitoring scheme to diagnose bearing distributed faults

  Most of the published work on the bearing condition monitoring via non-invasive techniques is focused on diagnosis of bearing localized defects. It is based on the some characteristic fault frequencies that appear in spectrum of the stator current or instantaneous power. Nonetheless, bearing distributed defects does not produce characteristic defect frequencies and thus could not been analysed via MCSA or IPA. These faults are common in industry, while they are often neglected in the research literature. As PVA performs the graphical analysis of stator current and does not require characteristic defect frequencies information for fault detection thus its capability to diagnose bearing distributed defects should be further researched.

- Development of hand-held instrument for on-site machine condition monitoring

  The IPA and PVA algorithms could be embedded to develop a hand-held instrument for on-site condition monitoring of induction motors. This will help to reduce the installation cost of the data acquisition system especially for small and medium scale industrial use.

- Reliability tests for mainstream industries

  The developed IPA and PVA algorithms should be expanded and utilized to analyse faults in turbines and generators. For consideration to be acceptable by the main stream industries, a reliability test of the developed algorithms should be conducted on large size motors (more than 40 HP).

- Multiple bearing faults analysis

  Even though the two techniques as proposed i.e. the instantaneous power analysis and Park vector analysis can monitor the conditions of induction motors from various scenarios, however these are not sufficient for most applications because outcomes of the analysis are based on the assumptions that each fault occurs independently. The extension of the approach as proposed in this work, to understand how each approach reacts to a combination of several faults (e.g. stator, rotor, eccentricity, bearings, etc.), would be useful.

12. Conclusion

The ability to forecast motor faults at incipient stages is vital to reducing maintenance costs, operation down-time and safety hazards. This paper synthesized the progress in the research

and development in condition monitoring and fault diagnosis of induction motors. The review presented in this paper has contributed to an improved understanding of the development of practical non-invasive condition monitoring and diagnostic tool for induction motors. The non-invasive methods, namely MCSA, PVA and IPA overcome the disadvantages associated to invasive methods. The MCSA, PVA and IPA can successfully diagnose the gear and bearing localized defects. However, further research is required to analyse bearing distributed faults via non-invasive condition monitoring methods. It has been observed that IPA has some implementation advantages over MCSA for incipient fault detection but IPA can be affected by environment noise perturbations. There is much scope for research to improve the capabilities of IPA in reliable online fault diagnosis system operating in an arbitrary environment noise. Furthermore, to realize the greatest economic and implementation benefits of IPA and PVA, it is important to design a hand-held condition monitoring system based on IPA and PVA techniques by considering the asset management perspective.

Acknowledgements

The authors acknowledge the support from Universiti Teknologi PETRONAS for providing Universiti Research Innovation Fund (URIF) and the Ministry of Higher Education (MOHE) Malaysia for the award of the Prototype Research Grant Scheme (PRGS).

Author details

Muhammad Irfan[1]*, Nordin Saad[2], Rosdiazli Ibrahim[2], Vijanth S. Asirvadam[2], Abdullah S. Alwadie[1] and Muhammad Aman Sheikh[2]

*Address all correspondence to: irfan16.uetian@gmail.com

1 Faculty of Engineering, Electrical Engineering Department, Najran University, Kingdom of Saudi Arabia

2 Electrical and Electronics Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Tronoh, Perak, Malaysia

References

[1] Irfan M, Saad N, Ibrahim R, Asirvadam VS, Magzoub M. An online fault diagnosis system for induction motors via instantaneous power analysis. Tribology Transactions. 2016.

[2] Irfan M, Saad N, Ibrahim R, Asirvadam VS, Hung NT, Magzoub MA. Analysis of bearing surface roughness defects in induction motors. Journal of Failure Analysis and Prevention. October 2015;**15**(5):730–736

[3] Irfan M, Saad, Ibrahim R, Asirvadam VS, Magzoub M. An intelligent fault diagnosis of induction motors in an arbitrary noisy environment. Journal of Nondestructive Evaluation. March 2016;**35**(12)

[4] Irfan M, Saad N, Ibrahim R, Asirvadam VS, Hung NT. Analysis of bearing outer race defects in induction motors. 5th International Conference on Intelligent and Advanced Systems (ICIAS); 2014

[5] Kia H, Henao SH, Capolino GA. Trends in gear fault detection using electrical signature analysis in induction machine-based systems. IEEE Workshop on Electrical Machines Design Control and Diagnosis (WEMDCD); 2015

[6] Sen PC. Principles of Electric machines and Power Electronics. John Wiley and Sons; 1989

[7] Irfan M, Saad N, Ibrahim R, Asirvadam VS. Development of an intelligent condition monitoring system for AC induction motors using PLC. IEEE Business Engineering and Industrial Applications Colloquium (BEIAC); 2013.

[8] Pineda M, Puche R, Riera M. Motor condition monitoring of induction motor with programmable logic controller and industrial network. IEEE Proceedings on Power Electronics; September 2011.

[9] IAS Motor Reliability Working Group. Report of large motor reliability survey of industrial and commercial installations—Part I. IEEE Transactions on Industry Applications. July/Aug 1985;**IA-21**:853–864

[10] IAS Motor Reliability Working Group, Report of large motor reliability survey of industrial and commercial installations—Part II. IEEE Transactions on Industry Applications. July/Aug 1985;**IA-21**:865–872

[11] IAS Motor Reliability Working Group. Report of large motor reliability survey of industrial and commercial installations—Part III. IEEE Transactions on Industry Applications. Jan. /Feb. 1987;**IA-23**:153–158

[12] Improved motors for Utility applications. Vol. 5. EPRI Publication EL-2678; October 2005

[13] Vas P. Parameter Estimation, Condition Monitoring, and Diagnosis of Electrical Machines. Oxford: Clarendon Press; 1993

[14] Tavner PJ, Ran L, Pennman J, Sedding H. Condition Monitoring of Rotating Electrical Machines. Letchworth, England: Research Studies Press Ltd.; 2008

[15] Fatigue and Fracture, ASM Handbook. Vol. 19. ASM International; 1996.

[16] Irfan M, Saad N, Ibrahim R, Asirvadam VS. A non-invasive method for condition monitoring of induction motors operating under arbitrary loading conditions. Arabian Journal for Science & Engineering. September 2016;**41**(9):3463–3471

[17] Suresh S. Fatigue of Materials. Cambridge, UK: Cambridge University Press; 1998

[18] Tallian TE. Failure Atlas for Hertz Contact Machine Elements. New York: ASME Press; 1992

[19] Olver AV. The mechanism of rolling contact fatigue: an update. Proceedings of the Institution of Mechanical Engineers. Part J: Journal of Engineering Tribology. 2005; **219**(5):313–330

[20] Halme J, Anderson P. Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics—state of the art. Proceedings of the Institution of Mechanical Engineers. Part J: Journal of Engineering Tribology. 2009;**224**(4)377–393

[21] Sadeghi F, Arakere NK, Jalalahmadi B, Slack TS, Raje N. A review of rolling contact fatigue. Journal of Tribology. 2009;**131**(4)

[22] Littmann WE. The Mechanism of Contact Fatigue. In: Ku PM, editor. Interdisciplinary Approach to the Lubrication of Concentrated Contacts. NASA SP-237. Washington, DC: National Aeronautics and Space Administration; 1970. pp. 309–378

[23] Littmann WE, Widner RL. Propagation of contact fatigue from surface and subsurface origins. Journal of Basic Engineering. 1966; **88**(3):624–636

[24] Nandi S, Toliyat HA. Condition monitoring and fault diagnosis of electrical machines—a review. 34th Annual Meeting of the IEEE Industry Applications. 1999. pp. 197–204

[25] Schoen RR, Habetler TG, Kamran F, Bartheld RG. Motor bearing damage detection using staor current monitoring. IEEE Transactions on Industry Applications. 1995;**31**(6): 1274–1279

[26] Eschmann P, Hasbargen L, Weigand K. Ball and Roller Bearings: Their Theory, Design, and Application. London: KG Heyden; 1958

[27] Dorrell DG, Thomson WT, Roach S. Analysis of air-gap flux, current, and vibration signals as function of a combination of static and dynamic eccentricity in 3-phase induction motors. IEEE Transactions on Industry Applications. 1997;**33**:24–34

[28] Wu S, Chow TWS. Induction machine fault detection using SOM-based RBF neural networks. IEEE Transactions on Industrial Electronics. February 2004;**51**(1),183–194

[29] Bradford M. Unbalanced magnetic pull in a 6-pole induction motor. IEEE Proceedings. 1968;**115**(11):1619–1627

[30] Dalpiaz G, Meneghetti U. Monitoring fatigue cracks in gears. NDT & E International. 1991;**24**(6):303–306

[31] Sottile J, Kohler JL. An on-line method to detect incipient failure of turn insulation in random-wound motors. IEEE Transactions on Energy Conversion. December, 1993;**8**(4):762–768

[32] Sheikh MA, Nor NM, Ibrahim T. A new method for detection of unbalance voltage supply in three phase induction motor. Jurnal Teknologi. 2016;**78**(5)

[33] Sheikh MA, Nor NM, Ibrahim T. Detection of unbalance voltage supply through rotor harmonics symbolic dynamics. The 6th International Conference on Intelligent & Advanced Systems, August 2016.

[34] Farag SF, Bartheld RG, May WE. Electronically enhanced low voltage motor protection and control. IEEE Transactions on Industry Applications. 1994;**29**(1):45–51

[35] Lee SB, and. Habetler TG. An online stator winding resistance estimation technique for temperature monitoring of line-connected induction machines. IEEE Transactions on Industry Applications. May 2003;**39**(3):685–694

[36] Bonnet AH, Soukup GC. Cause and analysis of stator and rotor failures in three-phase squirrel case induction motors. IEEE Transactions on Industry Application. July/August 1992;**28**(4)921–937

[37] Roux WL, Harley RG, Habetler TG. Detecting rotor faults in permanent magnet synchronous machines. SDEMPED'03; August 2003; Atlanta, GA; pp. 198–203

[38] Stack JR, Habetler TG, Harley RG. Bearing fault detection via autoregressive stator current modeling. IEEE Transactions Industry Applications. May 2004;**40**(3):740–747

[39] Cash MA. Detection of turn faults arising from insulation failure in the stator windings of AC machines [thesis]. USA: Georgia Institute of Technology; 1998

[40] Kankar PK, Sharma SC, Harsha SP. Nonlinear vibration signature analysis of a high speed rotor bearing system due to race imperfection. Journal of Computational and Nonlinear Dynamics. January 2012;**7**

[41] Babu CK, Tandon N, Pandey RK. Vibration modeling of a rigid rotor supported on the lubricated angular contact ball bearings considering six degrees of freedom and waviness on balls and races. Journal of Vibration and Acoustics. February 2012;134

[42] Immovilli F, Cocconcelli M, Bellini A, Rubini R. Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals. IEEE Transactions on Industrial Electronics. November 2009; **56**(11)

[43] Dolenc B, Boškoski P, Juričić D. Distributed bearing fault diagnosis based on vibration analysis. Mechanical Systems and Signal Processing. 2015;**66-67**:521–532

[44] Petersen D, Howard C, Sawalhi N, Ahmadi AM, Singh S. Analysis of bearing stiffness variations, contact forces and vibrations in radially loaded double row rolling element bearings with raceway defects. Mechanical Systems and Signal Processing. 2015;**50**:139–160

[45] Petersen D, Howard C, Prime Z. Varying stiffness and load distributions in defective ball bearings: Analytical formulation and application to defect size estimation. Journal of Sound and Vibration. 2015;**337**:284–300

[46] El-Thalji I, Antunen EJ. Fault analysis of the wear fault development in rolling bearings. Engineering Failure Analysis. 2015;**57**:470–482

[47] Prieto MD, Cirrincione G, Espinosa AG, Ortega JA, Henao H. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. IEEE Transactions on Industrial Electronics. August 2013;**30**(8)

[48] Jin X, Zhao M, Tommy WSC, Pecht M. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. IEEE Transactions on Industrial Electronics. May 2014;**61**(5)

[49] Shah DS, Patel VN. A review of dynamic modeling and fault identifications methods for rolling element bearing. Procedia Technology. 2014;**14**:447–456

[50] Singh S, Howard CQ, Hansen CH. An extensive review of vibration modelling of rolling element bearings with localised and extended defects. Journal of Sound and Vibration. June 2015

[51] Tanver PJ. Review of condition monitoring of rotating electrical machines. IET Electric Power Applications. 2008;**02**(4):215–247

[52] Zhou W, Habetler TG. Bearing condition monitoring methods for electric machines. IEEE Transaction on Diagnosis for Electric Machines, Power Electronics and Drives, 2007.

[53] Schoen RR, Lin BK, Habetler TG, Schlag JH, Farag S. An unsupervised, online system for induction motor fault detection using stator current monitoring. IEEE Transactions on Industry Applications. 1995;**31**(6):1280–1286

[54] Schoen RR, Habetler TG. Evaluation and implementation of a system to eliminate arbitrary load effects in current-based monitoring of induction machines. IEEE Transactions on Industry Applications. 1997;**33**(6):1571–1577

[55] Benbouzid MEH, Nejjari H, Beguenane R, Vieira M. Induction motor asymmetrical faults detection using advanced signal processing techniques. IEEE Transactions on Energy Conversion. June 1999;**14**(2)147–152

[56] Benbouzidi MEH, Viera M, Theys C. Induction motors faults detection and localization using stator current advanced signal processing techniques. IEEE Transactions on Power Electronics. January 1999;**14**(1):14–22

[57] Benbouzid MEH. A review of induction motors signature analysis as a medium for faults detection. IEEE Transaction on Industry Electronics. 2000;**47**(5)984–993

[58] Duque O, Pérez M, Moríñigo D. Detection of bearing faults in cage induction motors fed by frequency converter using spectral analysis of line current. IEEE Conference on Electric Machine Drives. May 2005;17–22

[59] Stack JR, Habetler TG, Harley RG. Bearing fault detection via autoregressive stator current modeling. IEEE Transactions on Industry Applications. May/Jun. 2004;**40**(3):740–747

[60] Bellini A, Immovilli F, Rubini R, Tassoni C. Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. IEEE IAS Annual Meeting. Edmonton, Alberta, Canada; October 2008; pp. 1–8.

[61] Zhou W, Habetler TG, Harley RG. Bearing fault detection via stator current noise cancellation and statistical contro. IEEE Transaction on Industrial Electronics. December 2008;**55**(12):4260–4269

[62] Stack JR, Habetler TG, Harley RG. Bearing fault detection via stator current modeling. IEEE Transactions on Industry Applications. 2004;**40**(3):740–747

[63] Liling S, Boqiang X. An improvement of stator current based detection of bearing fault in induction motors. IEEE Conference Record, Industry Applications Conference and 42nd IAS Annual Meeting 2007.

[64] Boqiang X, Huihuan Z, Liling S, Junzhong S. Weak-signal detection and the application in detection of electric motor faults. In: Proceeding of International Conference on Electrical Machines and Systems; Seoul, Korea. Oct. 8-11, 2007

[65] Trajin B, Regnier J, Faucher J. Indicator for bearing fault detection in asynchronous motors using stator current spectral analysis. IEEE International Symposium on Industrial Electronics; 2008.

[66] Blodt M, Granjon P, Raison B, Rostaing G. Models for bearing damage detection in induction motors using stator current monitoring. IEEE Transaction on Industrial Electronics. April 2008;**55**(4)

[67] Terra JI, Castelli M, Fossati JP. Fault detection and remote monitoring system for induction motors using MCSA technique. IEEE Transaction on Transmission and Distribution. October 2008;**11**(2)

[68] Bayindir R, Sfa I, Colak I. Fault detection and protection of induction motors using sensors. IEEE Transaction on Energy Conversion, September 2008;**23**(3)

[69] Ioannides MG. Design and implementation of PLC-based monitoring control system for induction motor. IEEE Transaction on Energy Conversion. September 2004;**19**(3)

[70] Frosini L, Bassi E. Stator current and motor efficiency as indicators for different types of bearing faults in induction motors. IEEE Transactions on Industrial Electronics. 2010;**57**(1):244–251

[71] Frosini L, Bassi E, Fazzi A, Gazzaniga C. Use of the stator current for condition monitoring of bearings in induction motors. ICEM; September 2008; Vilamoura, Portugal; 1–6

[72] Ebrahimi BM, Faiz J. Feature extraction for short-circuit fault detection in permanent-magnet synchronous motors using stator-current monitoring. IEEE Transactions on Power Electronics. October 2010;**25**(10)

[73] Mehala N. Condition Monitoring and Fault Diagnosis of Induction Motor using Motor Current Signature Analysis [PhD Thesis]. Kurukshetra, India: National Institute of Technology; 2010.

[74] Romero-Troncoso RJ, Saucedo-Gallaga R, Cabal-Yepez E, Garcia-Perez A, Osornio-Rios RA. FPGA-based online detection of multiple combined faults in induction motors through information entropy and fuzzy inference. IEEE Transactions on Industrial Electronics. November 2011;**58**(11)

[75] Seera M, Lim CP, Ishak D, Singh H. Fault detection and diagnosis of induction motors using motor current signature analysis and a hybrid FMM–CART model. IEEE Transactions on Neural Networks and Learning Systems. January 2012;**23**(1)

[76] Singh S, Kumar A, Kumar N. Motor current signature analysis for bearing fault detection in mechanical systems. Procedia Materials Science. 2014;**6**:171–177

[77] Esfahani ET, Wang S, Sundararajan V. Multisensor wireless system for eccentricity and bearing fault detection in induction motors. IEEE/ASME Transactions on Mechatronics. June 2014;**19**(3)

[78] Choi S, Pazouki E, Baek J, Bahrami HR. Iterative condition monitoring and fault diagnosis scheme of electric motor for harsh industrial application. IEEE Transactions on Industrial Electronics. 2014. DOI: 10.1109/TIE.2014.2361112

[79] Shi P, Chen Z, Vagapov Y. Modelling and analysis of induction machines under broken rotor-bar failures. International Journal of Computer Applications. May 2013;**69**(14)

[80] Choi S, Akin B, Kwak S, Toliyat HA. A compact error management algorithm to minimize false-alarm rate of motor/generator faults in (hybrid) electric vehicles. IEEE Journal of Emerging and Selected Topics in Power Electronics. 2014. DOI: 10.1109/JESTPE.2014.2302902

[81] Choi SD, Akin B, Rahimian M, Toliyat HA. Implementation of fault diagnosis algorithm for induction machines based on advanced digital signal processing techniques. IEEE Transaction on Industrial Electronics. March 2011;**58**(3)

[82] Aydin I, Karakose M, Akin E. Combined intelligent methods based on wireless sensor networks for condition monitoring and fault diagnosis. Journal of Intelligent Manufacturing. August 2013. DOI: 10.1007/s10845-013-0829-8

[83] Blodt M, Granjon P, Raison B, Rostaing G. Models for bearing damage detection in induction motors using stator current monitoring. IEEE Transaction on Industrial Electronics. April 2008;**55**(4)

[84] Stack JR, Habetler TG, Harley RG. Fault classification and fault signature production for rolling element bearing in electric machines. IEEE Transaction on Industry Applications. 2004;**40**(3)

[85] Benbouzid MEH, Nejjari H, Beguenane R, Vieira M. Induction motor asymmetrical faults detection using advanced signal processing techniques. IEEE Transaction on Energy Conversion. June 1999;**14**(2):147–152

[86] Patel VN, Tandon N, Pandey RK. Defect detection in deep groove ball bearing in presence of external vibration using envelope analysis and duffing oscillator. Measurement. 2012;**45**:960–970

[87] Kang M, Kim J, Kim JM. Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm. Information Sciences. 2015; **294**:423–438

[88] Choi SD, Akin B, Rahimian MM, Toliyat HA, Azadpour M. A generalized condition monitoring method for multi-phase induction motors. IEEE International Conference on Electric Machines and Drives; 2009

[89] Wang D, Miao Q, Fan X, Huang HZ. Rolling element bearing fault detection using an improved combination of Hilbert and Wavelet transforms. Journal of Mechanical Science and Technology. August 2009;**23**:3292–3301

[90] Widodo A, Yang BS, Han T. Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. Expert Systems with Applications. February 2007;**32**(2):299-312

[91] Hwang Y, Jen K, Shen Y. Application of cepstrum and neural network to bearing fault detection. Journal of Mechanical Science and Technology. May 2009;**23**:2730-2737

[92] Júnior AMG, Silva VR, Baccarini LMR, Reis MLF. Three-phase induction motors faults recognition and classification using neural networks and response surface models. Journal of Control, Automation and Electrical Systems. June 2014;**25**(3):330-338

[93] Kia SH, Henao H, Capolino GA. Development of a test bench dedicated to condition monitoring of wind turbines. IEEE-IECON; Dallas (TX, USA); October 28–November 1; 2014

[94] Guzinski J, Diguet M, Krzeminski Z, Lewicki A, Abu-Rub H. Application of speed and load torque observers in high speed train drive for diagnostic purpose. IEEE Transactions on Industrial Electronics. January 2009;**56**(1):248-256

[95] Mohanty AR, Kar C. Monitoring gear vibrations through motor current signature analysis and wavelet transform. Mechanical System and Signal Processing. January 2006;**20**(1):158-187

[96] Feki N, Clerc G, Velex P. An integrated electro-mechanical model of motor-gear units - Applications to tooth fault detection by electric measurements. Mechanical System and Signal Processing. May 2012;**29**:377-390

[97] Strangas EG. Response of electrical drives to gear and bearing faults—diagnosis under transient and steady state conditions. In: Proceedings of Workshop on Electrical Machines Design Control and Diagnosis (WEMDCD); invited paper, March 11–12, 2013; Paris (France), pp.289-297

[98] Henao H, Kia SH, Capolino GA. Torsional vibration assessment and gear fault diagnosis in railway traction system. IEEE Transactions on Industrial Electronics. May 2011;**58**(5):1707-1717

[99] Bogiatzidis I, Safacas A, Mitronikas E. Detection of backlash phenomena appearing in a single cement kiln drive using the current and the electromagnetic torque signature. IEEE Transactions on Industrial Electronics. August 2013;**60**(8):3441-3453

[100] Trajin B, Regnier J, Faucher J. Comparison between stator current and estimated mechanical speed for the detection of bearing wear in asynchronous drives. IEEE Transactions on Industrial Electronics. November 2009;**56**(11):4700-4709

[101] Ibrahim A, El-Badaoui M, Guillet F, Bonnardot F. A new bearing fault detection method in induction machines based on instantaneous power factor. IEEE Transactions on Industrial Electronics. December 2008;55(12):4252–4259

[102] Blödt M, Chabert M, Regnier J, Faucher J. Mechanical load fault detection in induction motors by stator current time-frequency analysis. IEEE Transactions on Industry Applications. November 2006; **42**(6):1454–1463

[103] Mohanty AR, Kar C. Fault detection in a multistage gearbox by demodulation of motor current waveform. IEEE Transactions on Industrial Electronics. June 2006;**53**(4): 1285–1297

[104] Kar C, Mohanty AR. Vibration and current transient monitoring for gearbox fault detection using multiresolution Fourier transform. Journal of Sound and Vibration. March 2008; **311**(1/2):109–132

[105] Kia SH, Henao H, Capolino GA. Analytical and experimental study of gearbox mechanical effect on the induction machine stator current signature. IEEE Transactions on Industrial Electronics. July 2009;**45**(4):1405–1415

[106] Kia SH, Henao H, Capolino GA. Gear tooth surface damage fault detection using induction machine stator current space vector analysis. IEEE Transactions on Industrial Electronics. March 2015;**62**(3):1866–2001

[107] Kia SH, Henao H, Capolino GA. A comparative study of acoustic, vibration and stator current signatures for gear tooth fault diagnosis. International Conference on Electrical Machines; 2-5 September 2012; Marseille.

[108] Kia SH, Henao H, Capolino GA. Gear tooth surface damage fault detection using induction machine electrical signature analysis. 9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED); August 27–30, 2013;Valencia.

[109] Kia SH, Henao H, Capolino GA. Trends in gear fault detection using electrical signature analysis in induction machine-based systems. IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD); March 26-27, 2015; Torino.

[110] Kia SH, Henao H, Capolino GA. Torsional vibration effects on induction machine current and torque signatures in gearbox-based electromechanical system. IEEE Transactions on Industrial Electronics. November 2009;**56**(11):4689–4699

[111] Feki N, Clerc G, Velex P. Gear and motor fault modeling and detection based on motor current analysis. Electric Power System Research. Feburary 2013;95:28–37

[112] Daneshi-Far Z, Henao H, Capolino GA. Planetary gearbox effects on induction machine in wind turbine: Modeling and analysis. In: Proceedings of International Conference on Electrical Machines (ICEM); 2–5 September 2012; Marseille (France); pp.1790–1796

[113] Ottewill JR, Orkisz M. Condition monitoring of gearboxes using synchronously averaged electric motor signals, Mechanical System Signal Processing. July 2013;**38**(2): 482–498

[114] Girsang IP, Dhupia JS, Muljadi E, Singh M, Pao LY. Gearbox and drive train models to study of dynamic effects of modern wind turbines. IEEE Transactions on Industry Applications. 2014

[115] Huh KK, Lorenz RD, Nagel NJ. Gear fault diagnostics integrated in the motion servo drive for electromechanical actuators. IEEE Transactions on Industry Applications. July 2012;**48**(1):142–150

[116] Bogiatzidis IX, Safacas AN, Mitronikas ED. Detection of backlash phenomena appearing in a single cement kiln drive using the current and the electromagnetic torque signature. IEEE Transactions on Industrial Electronics. August 2013;**60**(8)

[117] Kia SH, Henao H, Capolino GA. A real-time platform dedicated to online gear tooth surface damage fault detection in induction machine. International Conference on Electrical Machines (ICEM); 2-5 September, 2014; Berlin.

[118] Fournier E et al. Current-based detection of mechanical unbalance in an induction machine using spectral kurtosis with reference. IEEE Transactions on Industrial Electronics, March 2015;**62**(3):1879–1887

[119] Ahmad I. Investigation of single and multiple faults under varying load conditions using multiple sensor types to improve condition monitoring of induction machines [PhD Thesis], Australia: Univeristy of Adelaide; 2007.

[120] Drif M, Marques Cardoso AJ. Stator fault diagnostics in squirrel cage three-phase induction motor drives using the instantaneous active and reactive power signature analyses. IEEE Transaction on Industrial Electronics. May 2014;**10**(2)

[121] Cruz SMA. An active–reactive power method for the diagnosis of rotor faults in three-phase induction motors operating under time-varying load conditions. IEEE Transactions on Energy Conversion. March 2012;**27**(1)

[122] Rjakishmi Samaga LR, Vittal KP. Comprehensive study of mixed eccentricity fault diagnosis in induction motors using signature analysis. Electrical Power and Energy Systems. 2012;**35**:180–185

[123] Kim J, Shin S, Lee SB, Gyftakis KN, Drif M, Marques Cardoso AJ. Power spectrum-based detection of induction motor rotor faults for immunity to false alarms. IEEE Transactions on Energy Conversion. 2015

[124] Didier G, Ternisien E, Caspary O, Razik H. Fault detection of broken rotor bars in induction motor using a global fault index. IEEE Transactions on Industry Applications. January 2006; **42**(1):79–88

[125] Zhou W, Habetler TG, Harley RG. Incipient bearing fault detection via motor stator current noise cancellation using Wiener filter. IEEE Transactions on Industry Applications. July 2009;45(4):1309–1317

[126] Golafshan R, Sanliturk KY. SVD and Hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults. Mechanical Systems and Signal Processing. 2015

[127] Golafshan R, Sanliturk KY. The Effect of SVD based noise elimination methods applied for ball bearing fault detection. TrC-IFToMM Symposium on Theory of Machines and Mechanisms; June 14–17, 2015; Izmir, Turkey.

[128] Kia SH, Henao H, Capolino G. A high-resolution frequency estimation method for three-phase induction machine fault detection. IEEE Transaction on Industrial Electronics. August 2007;**54**(4):2305–2314

[129] Bellini A, Yazidi A, Filippetti F, Rossi C, Capolino GA. High frequency resolution techniques for rotor fault detection of induction machines. IEEE Transactions on Industrial Electronics. December 2008;**55**(12):4200–4209

[130] Kim H, Lee SB, Park SB, Kia SH, Capolino GA. Reliable detection of rotor faults under the influence of low frequency load torque oscillations for applications with speed reduction couplings. IEEE 10th International Symposium on Diagnostics for Electrical Machines. Power Electronics and Drives; September, 2015; Portugal.

[131] Dong J, Chin J. Noise resistant time frequency analysis and application in fault diagnosis of rolling element bearings. Mechanical Systems and Signal Processing. 2012;**33**:212–236

[132] Wang Z, Jiang H. Robust incipient fault identification of aircraft engine rotor based on wavelet and fraction. Aerospace Science and Technology. 2010;**14**:221–224

[133] Akin B, Orguner U, Toliyat HA, Rayner M. Phase-sensitive detection of motor fault signatures in the presence of noise. IEEE Transactions on Industrial Electronics. June 2008;**55**(6):2539–2550

[134] Akin B, Choi SD, Orguner U, Toliyat HA. A simple real-time fault signature monitoring tool for motor drive imbedded fault diagnosis systems. IEEE Transaction on Industrial Electronics. May 2011;**58**(5)

[135] Toliyat HA, Nandi S, Choi S, Kelk HS. Electric machines, modelling, condition monitoring and fault diagnosis. CRC Press, Taylor & Francis Group; 2012.

[136] Rajagopalan S, Habetler TG, Harley RG, Restrepo JA, Alle JM. Non-stationary motor fault detection using recent quadratic time-frequency representations. International Conference Recording IEEE IAS Annual Meeting; October 2006; **5**; pp. 2333–2339

[137] Jung JH, Lee JJ, Kwon BH. Online diagnosis of induction motor using MCSA. IEEE Transactions on Industry Applications. December 2006;**53**(1):1842–1852

[138] Yousuf A, Delpha C, Diallo D. An optimal fault detection threshold for early detection using Kullback–Leibler divergence for unknown distribution data. Signal Processing. 2015

[139] Harmouche J, Delpha C, Diallo D. Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part II. Signal Processing. 2015;**109**:334–344

[140] Delpha C, Diallo D. Incipient fault detection and diagnosis: A hidden information detection Problem. IEEE 24th International Symposium on Industrial Electronics; June, 2015; Buzios.

[141] Chen J, Patton RJ. Robust Model Based Fault Diagnosis for Dynamic Systems. Springer Science. Business Media; 2012.

[142] Fournier E, Picot A, Régnierl J, Yamdeu MT. On the use of spectral kurtosis for diagnosis of electrical machines. The 9th IEEE International Symposium on diagnostics for electric machines, power electronics and drives (SDEMPED); 2013; Valencia

[143] Picot A, Obeid Z, Régnierl J, Poignant S, Darnis O, Maussion P. Statistic-based spectral indicator for bearing fault detection in permanent-magnet synchronous machines using the stator current. Mechanical Systems and Signal Processing. 2014;**46**:424–441

[144] Zarie J. Induction motors bearing fault detection using pattern recognition techniques. Expert System with Applications. 2012;**39**:68–73

[145] Spyropoulos DV, Mitronikas ED. Induction motor stator fault diagnosis technique using Park vector approach and complex Wavelets. The 20th IEEE International Conference on Electrical Machines (ICEM); September, 2012; Marseille.

[146] Laughman C, Leeb SB, Norford LK, Shaw SR, Armstrong PR. A Park transform-based method for condition monitoring of three-phase electromechanical systems. The 5th IET International Conference on Power Electronics, Machines and Drives; April 2010; Brighton, UK.

[147] Salem SB, Touti W, Bacha K, Chaari A. Induction motor mechanical fault identification using park vector approach. International Conference on Electrical Engineering and Software Applications (ICEESA); March 2013.

[148] Parra AP, Enciso AMC, Ochoa JO, Peñaranda JAP. Stator fault diagnosis on squirrel cage induction motors by ESA and EPV. IEEE Workshop on Power Electronics and Power Quality Applications (PEPQA); 2013; Bogota, July.

[149] Rezig A, N'Diaye A, Mekideche MR, Djerdir A. Modelling and detection of bearing faults in permanent magnet synchronous motors. The 20th IEEE International Conference on Electrical Machines (ICEM). September, 2012; Marseille.

[150] Bacha K, Salem SB, Chaari A. An improved combination of Hilbert and Park transforms for fault detection and identification in three-phase induction motors. Electrical Power and Energy Systems. 2012;**43**:1006–1016

[151] Salem SB, Bacha K, Gossa M. Induction motor fault diagnosis using an improved combination of Hilbert and Park transforms. The 16th IEEE Mediterranean Electrotechnical Conference (MELECON); March 2012; Yasmine Hammamet.

[152] Kuruppu SS, Kulatunga NA. D-Q current signature-based faulted phase localization for SM-PMAC machine drives. IEEE Transactions on Industrial Electronics. January 2015;**62**(1)

[153] Zhang P, Du Y, Habetler TG, Lu B. A survey of condition monitoring and protection methods for medium-voltage induction motors. IEEE Transactions on Industry Applications. January 2011;**47**(1):34–46

[154] Navarro L, Delgado M, Urresty J, Cusidó J, Romeral L. Condition monitoring system for characterization of electric motor ball bearings with distributed fault using fuzzy inference tools. IEEE Instrumentation and Measurement Technology Conference (I2MTC); May, 2010; Austin.

[155] Immovilli F, Cocconcelli M, Bellini A, Rubini R. Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals. IEEE Transactions on Industrial Electronics. November 2009;**56**(11):4710–4717

[156] Immovilli F, Bellini A, Rubini R, Tassoni C. Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. IEEE Transactions on Industry Applications. July 2010;**46**(4):1350–1359

[157] Dalvand F, Keshavarzi M, Kalantar A, Cheraghdar A. Detection of generalized-roughness bearing fault using statistical-time indices of instantaneous frequency of motor voltage space vector. The 23rd Iranian Conference on Electrical Engineering (ICEE); 2015.

[158] Gao Z, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques part I: fault diagnosis with model based and signal-based approaches. IEEE Transactions on Industrial Electronics. 2015.

[159] Hurtado ZYM, Tello CP, Sarduy JG. A review on detection and fault diagnosis in induction machines. Publicaciones en Ciencias y Tecnologa. July 2014;**8**(1)

[160] Heng A, Zhang S, Tan ACC, Mathew J. Rotating machinery prognostics: State of the art, challenges and opportunities—A review. Mechanical Systems and Signal Processing. 2009;**23**:724–739

[161] Zhang P, Du Y, Habetler TG, Lu B. A survey of condition monitoring and protection methods for medium-voltage induction motors. IEEE Transactions on Industry Applications. 2011;**47**(1)

[162] Gandhi A, Corrigan T, Parsa L. Recent advances in modeling and online detection of stator interturn faults in electrical motors. IEEE Transactions on Industrial Electronics. May 2011;**58**(5)

[163] Hameed Z, Hong YS, Cho YM, Ahn SH, Song CK. Condition monitoring and fault detection of wind turbines and related algorithms: A review. Renewable and Sustainable Energy Reviews. 2009;**13**:1–39

[164] Irfan M, Saad N, Ibrahim R, Asirvadam VS. An online condition monitoring system for induction motors via instantaneous power analysis. Journal of Mechanical Science and Technology. April 2015; **29**(04)

[165] Irfan M, Saad N, Ibrahim R, Asirvadam VS. Condition monitoring of induction motors via instantaneous power analysis. Journal of Intelligent Manufacturing. February 2015

[166] Irfan M, Saad N, Ibrahim R, Asirvadam VS. An intelligent diagnostic condition monitoring system for ac motors via instantaneous power analysis. International Review of Electrical Engineering. April 2013; **8**(2):664–672

[167] Irfan M, Saad N, Ibrahim R, Asirvadam VS.. An approach to diagnose inner race surface roughness faults in bearings of induction motors. IEEE International Conference on Signal and Image Analysis (ICSIPA); October 2015; Kuala Lumpur, Malaysia.

[168] Irfan M, Saad N, Ibrahim R, Asirvadam VS.. Diagnosis of distributed faults in outer race of bearings via Park's Transformation method. The 10th Asian Control Conference (ASCC); June 2015; Kota Kinabalu, Malaysia.

[169] Irfan M, Saad N, Ibrahim R, Asirvadam VS.. A non-invasive fault diagnosis system for induction motors in noisy environment. IEEE International Conference on Power and Energy (PECon); December 2014; Malaysia.

[170] Irfan M, Saad N, Ibrahim R, Asirvadam VS. An intelligent diagnostic system for condition monitoring of ac motors. The 8th IEEE Conference on Industrial Electronics and Applications; June 2013; Melbourne, Australia.

[171] Zarei J, Poshtan J. An advanced Park's vectors approach for bearing fault detection. Tribology International. 2009;**19**:213–219

Chapter 5

Evaluation of Novelty Detection Methods for Condition Monitoring applied to an Electromechanical System

Miguel Delgado Prieto, Jesús A. Cariño Corrales,
Daniel Zurita Millán, Marta Millán Gonzalvez,
Juan A. Ortega Redondo and
René de J. Romero Troncoso

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67531

Abstract

Dealing with industrial applications, the implementation of condition monitoring schemes must overcome a critical limitation, that is, the lack of a priori information about fault patterns of the system under analysis. Indeed, classical diagnosis schemes, in general, outdo the membership probability of a measure in regard to predefined operating scenarios. However, dealing with noncharacterized systems, the knowledge about faulty operating scenarios is limited and, consequently, the diagnosis performance is insufficient. In this context, the novelty detection framework plays an essential role for monitoring systems in which the information about different operating scenarios is initially unavailable or restricted. The novelty detection approach begins with the assumption that only data corresponding to the healthy operation of the system under analysis is available. Thus, the challenge is to detect and learn additional scenarios during the operation of the system in order to complement the information obtained by the diagnosis scheme. This work has two main objectives: first, the presentation of novelty detection as the current trend toward the new paradigm of industrial condition monitoring and, second, the introduction to its applicability by means of analyses of different novelty detection strategies over a real industrial system based on rotatory machinery.

Keywords: condition monitoring, electromechanical systems, failure diagnosis, feature reduction, industrial monitoring applications, novelty detection, open set recognition

1. Introduction

Currently, condition monitoring plays a key role in the reliability and safety strategies of most of the industrial applications [1]. Classical industrial condition monitoring methodologies imply the estimation of numerical features and their posterior processing in order to characterize the available physical magnitudes acquired during the operation of the system under analysis. Such numerical feature vectors are, then, presented to a classification algorithm in order to obtain a diagnosis outcome [2]. In this procedure, the algorithm of classification is previously trained with available data representative of different system conditions. Thus, during the regular operation of the condition monitoring scheme, each measurement acquired from the system will be transformed to a vector of numerical features, and its similarities with previous patterns will be evaluated in order to obtain the related probability. During the last decades, a great deal of studies has been done around different aspects of the electromechanical condition monitoring, that is, the potentiality of different physical magnitudes for fault detection, the analysis of time, frequency and time-frequency domains for numerical features' estimation, the effect of feature reduction techniques for patterns' characterization and dealing with data-based approaches and multiple classification strategies for diagnosis improvement [3, 4]. All of these works are, with no doubt, a major step forward to the study, research and development of enhanced condition monitoring schemes to be applied to electromechanical systems. However, currently, the scientific and industrial communities are working together toward more demanding industrial challenges in the frameworks of Industry 4.0 [5] and Zero-defect manufacturing [6]. Indeed, further capabilities are expected from the condition monitoring developments in order to face questions about their practical implementations, questions such as: *How must condition monitoring be managed in front of new operating scenarios not previously considered?*, *How to detect new operating scenarios?*, *Which numerical features should be used for unknown patterns' detection?*, *How to preserve the diagnosis reliability in the presence of new patterns?*, *Is it possible to automate these considerations or is the aid of an expert required?* In order to find answers to such questions, specific research is being gathered around the so-called novelty detection topic, which can be defined as the task of recognizing that the data under analysis differ, in some respect, from the initial available data.

Indeed, a priori characteristic fault patterns of specific rotatory machinery are not usually available and highly difficult to estimate through theoretical approaches. Thus, condition monitoring strategies capable of detecting novel operating conditions, alongside the classification of known conditions, represent the most convenient solutions [7]. This approach is known as the open set recognition problem, where only a reduced set of known operating scenarios are included in the initial dataset and used during the training stage, and, then, novel (unknown) scenarios may appear during the online diagnosis stage. In general terms, in order to deploy a novel detection strategy, a model must be trained with all the available data describing the initial-known scenarios of the machinery under monitoring. Thus, the model generates a threshold system that allows to discriminate between known scenarios and measurements corresponding to new cases, novelties. Different approaches differ in the way that the threshold system is generated [8].

http://dx.doi.org/10.5772/67531

In this chapter, first, a more comprehensive description of the novelty detection topic, including different approaches and their dependencies, is introduced. Later, the practical application of novelty detection applied over an industrial electromechanical system is described. The performances obtained with different novelty detection strategies, including the effect of feature reduction, are discussed finally.

2. Novelty detection

The introduction of novelty detection into the classical monitoring chain represents a previous condition to the diagnosis assessment. The classical step flow to implement novelty detection is shown in **Figure 1**. The procedure begins with the off-line processing of the available information (generally, the healthy behavior of the electromechanical chain). Such processing, in regard to the raw data acquired (stator currents, temperatures, etc.), consists of the definition of the same blocks as in classical diagnosis procedures, that is, feature estimation (calculation of a set of numerical features) and feature reduction (feature vector transformation for improved characterization). Once the available data is characterized by vectors of D features, the configuration of the novelty detection model follows. This part depends entirely on the nature of the novelty model (different approaches can be applied); however, the objective is the

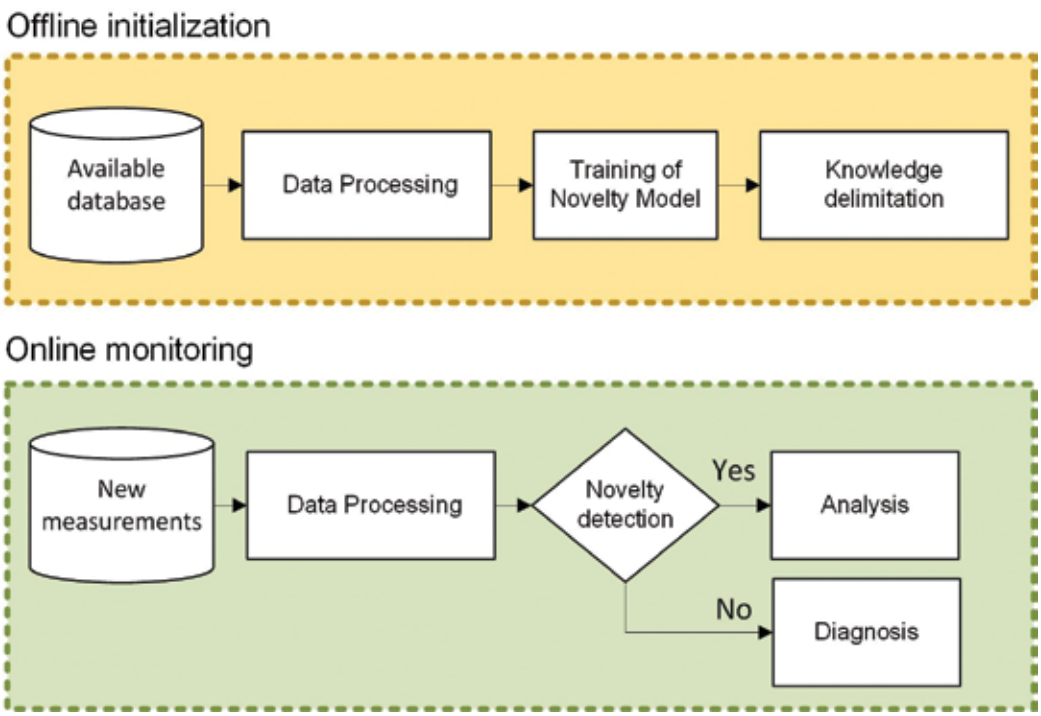


Figure 1. Implementation scheme of novelty detection, including the offline initialization for knowledge delimitation and the online monitoring for novelty evaluation.

delimitation of the available knowledge defining a set of mathematical descriptions in the *D*-dimensional feature space in which the available database is projected [9].

Thus, during the online monitoring, the novelty detection model will analyze the new acquisitions and will determine if the new data correspond to known operating scenarios previously learned or present different characteristics and can be considered novel representations. In case of known operating scenarios, the diagnosis follows. It must be noted that the diagnosis procedure could include different feature estimation and feature reduction stages because its objective is completely different. In this sense, the novelty detection does not require different labels since all data belongs to the class *knowledge* or *normal*. The diagnosis, however, requires to maintain different labels in order to allow the identification of the different operating scenarios. In case of unknown operating scenarios, the diagnosis cannot be carried out since the diagnosis reliability would be affected. In this case, the presence of novel data is reported and, after the supervision of an expert, the measurements are stored in order to upgrade the known operating scenarios, which will imply the retraining of the novelty model [10].

In order to illustrate the novelty detection operation, an example is shown in **Figure 2**. A D = 2-dimensional feature space is considered, in which a set of measurements representing the available data has led to the definition of the boundaries corresponding to the known conditions. When new measurements are acquired, the novelty model analyzes them and determines, in this example by means of their position in the 2-dimensional feature space, if they represent novelty or if the behavior is still considered known. If a significant amount of novel acquisitions with the same characteristics are detected, then, a novel operation mode is detected and, if validated, the data will be included as known behaviors.

Indeed, the detection of novel events is an important ability of any condition monitoring scheme. Considering the fact that it cannot be trained within a machine learning system with all possible systems' variability, it becomes important to include the differentiation capability between *known* and *unknown* object information during the system's monitoring. However, it has been considered in practice by several studies that the novelty detection is an extremely

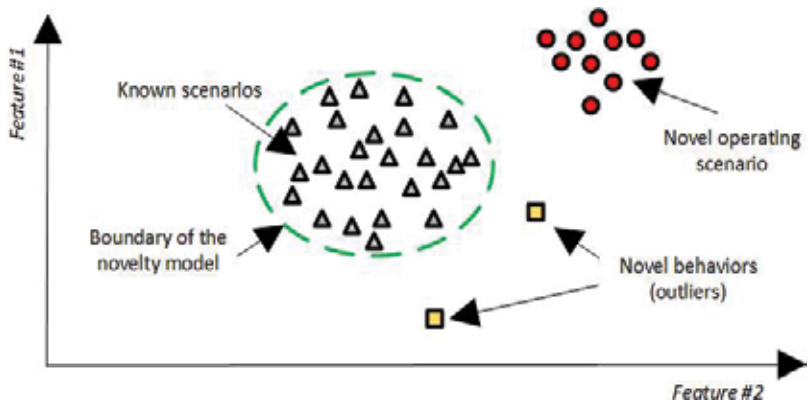


Figure 2. Example of a novelty detection basic operation, including available data, Δ, mathematical description of the available data, --, novel behaviors detected as outliers, □ and novel behaviors detected as the novel operating scenario, O.

http://dx.doi.org/10.5772/67531

challenging task. It is for this reason that there exist different approaches of novelty detection that have been demonstrated to perform well under different applications [11, 12]. Unfortunately, it is clearly evident that there is no single best model for novelty detection, and the success depends not only on the type of the method used but also on the statistical properties of the available data. Next, the three basic novelty detection approaches are described, including probabilistic, domain-based and distance-based methods.

2.1. Probabilistic methods

Probabilistic approaches to novelty detection are based on estimating the probability density function (PDF) of the available data. The resulting distribution may then be thresholded to define the boundaries of *normality* in the feature space and assess whether a new measurement belongs to the same distribution or not. The training data is assumed to be generated from some underlying probability distribution. This estimation usually represents the novelty model, and a novelty threshold can be set over such estimation. The estimation of the underlying data density from a multivariate training dataset is a well-established topic [10].

Probabilistic methods are divided in parametric and nonparametric approaches. Parametric approaches impose a restrictive model on the data, which results in a large bias when the model does not fit the data. Nonparametric approaches set up a very flexible model by making fewer assumptions over the data: The model grows in size to accommodate the complexity of the data, but it requires a large sample size for a reliable fit out of all the free parameters. The opinion in the scientific literature is divided as to whether various techniques should be classified as parametric or nonparametric. For the purposes of providing probabilistic estimators, Gaussian Mixture Model (GMM) and Kernel Density Estimator (KDE) have proven popular. The GMM is typically classified as a parametric technique [11] because of the assumption that the data is generated from a weighted mixture of Gaussian distributions. The KDE is typically classified as a nonparametric technique [13] as it is closely related to histogram methods, one of the earliest forms of nonparametric density estimation approaches.

2.1.1. Gaussian-mixture model

The GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. The GMM parameters are estimated from the available training data using, for example, the iterative expectation-maximization algorithm or the maximum a posteriori estimation. Thus, a GMM is a weighted sum of M component Gaussian densities, mathematically described as,

$$p(x|\lambda) = \sum_{i=1}^{M} w_i g\left(x|\mu_i, \sum i\right) \tag{1}$$

where x is a D-dimensional vector, w_i, i=1,..M are the mixture weights and $g(x|u_i, \sum i)$, i=1,..,M are the component Gaussian densities. Each component density is a D-variate Gaussian function of the form,

$$g(x|\mu_i, \sum i) = \frac{1}{(2\pi)^{D/2}|\sum i|^{1/2}} exp\left\{ -\frac{1}{2}(x-\mu_i)' \sum_{i}^{-1}(x-\mu_i) \right\} \quad (2)$$

with mean vector μ_i and covariance matrix Σ_i. The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$. The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation $\lambda = \{w_i, \mu_i, \Sigma_i\},\ i = 1, .., M$. There are several variants on the GMM. The covariance matrices, Σ_i, can be in full rank or constrained to be diagonal. Additionally, parameters can be shared, or tied, among the Gaussian components, such as having a common covariance matrix for all components. The choice of model configuration is often determined by the amount of data available for estimating the GMM parameters and how the GMM is used in a particular application [14]. In fact, GMM can suffer from the requirement of large numbers of training examples to estimate model parameters. A further limitation of parametric techniques is that the chosen functional form for the data distribution may not be a good model of the distribution that generates the data.

One of the major issues in novelty detection is the selection of a suitable novelty threshold. Within a probabilistic approach, novelty scores can be defined using the unconditional probability distribution $z(x) = p(x)$ and a typical approach to setting a novelty threshold k is to threshold this value, that is, $p(x) = k$. However, because $p(x)$ is a probability density function, a threshold on $p(x)$ has no direct probabilistic interpretation. Some studies have interpreted the model output $p(x)$ probabilistically, by considering the cumulative probability P associated with $p(x)$, that is, determining the probability mass obtained by numerically estimating the integral of $p(x)$ over a region R for which the value of $p(x)$ is above the novelty threshold k [15]. For unimodal distributions, one can integrate from the mode of the probability density function to the probability contour defined by the novelty threshold $p(x) = k$, which can be achieved in a closed form for most regular distributions.

2.1.2. Kernel density estimator

Nonparametric approaches do not assume that the structure of a model is fixed, that is, the model grows in size necessary to fit the data and accommodates the complexity of the data. The simplest nonparametric statistical technique is the use of histograms. The algorithm typically defines a distance measure between a new test data point and the histogram-based model of normality to determine if it is an outlier or not [10]. For multivariate data, attribute-wise histograms are constructed and an overall novelty score for a test data point is obtained by aggregating the novelty scores from each attribute. However, when a histogram is defined, it is necessary to consider the width of the bins (equal subintervals in which the whole data interval is divided) and the end points of the bins (where each of the bins starts). In consequence, the histograms present a nonsmooth behavior. In order to alleviate this deficiency, the kernel estimators were proposed.

It must be considered that observations are being drawn from some unknown probability density function $p(x)$ in a Euclidian D-dimensional feature space. Thus, considering a region R containing the D-dimensional measurement x, the probability mass associated with this

http://dx.doi.org/10.5772/67531

region is given by $P = \int_R p(x)dx$. Taking into account a dataset comprising N observations drawn from $p(x)$, each data point has a probability P of falling within R, and the total number K of points that lie inside R will be distributed according to the binomial distribution $Bin(K|N,P) = \frac{N!}{K!(N-K)!}P^K(1-P)^{1-k}$. The mean fraction of points falling inside the region is $E[K|N] = P$, and the variance around this mean is $var[K|N] = P(1-P)/N$. For large N, this distribution will sharply peak around the mean and so $K \cong NP$. If, however, it is assumed that the region R is sufficiently small that the probability density $p(x)$ is roughly constant over the region, then $P \cong p(x)V$, where V is the volume of R. Thus, density estimate is obtained in the form,

$$P(x) = \frac{K}{NV} \tag{3}$$

Note that the validity of Eq. (3) depends on two contradictory assumptions, namely that the region R is sufficiently small that the density is approximately constant over the region and yet sufficiently large (in relation to the value of that density) that the number K of points falling inside the region is sufficient for the binomial distribution to sharply peak. The resultant Eq. (3) can be exploited in two different ways. Either it can be fixed K and the value of V can be determined from the data, which gives rise to the K-nearest-neighbor technique that will be presented later, or it can be fixed V and K can be determined from the data, giving rise to the kernel approach. It can be shown that both the K-nearest-neighbor density estimator and the kernel density estimator converge to the true probability density in the limit $N \to \infty$, provided V shrinks suitably with N, and K grows with N [13]. Thus, considering the region R as a small hypercube centered on the point x at which is desired to determine the probability density, the number K of points falling within region is defined as follows,

$$K(u) = \begin{cases} 1, & |u_i| \leq \frac{1}{2}, \ i = 1,..,D \\ 0, & otherwise \end{cases} \tag{4}$$

which represents a unit cube centered on the origin. The function $K(u)$ is an example of a kernel function and in this context is also called a Parzen window. From Eq. (4), the quantity $K\left(\frac{x-x_i}{h}\right)$ will be one if the data point x_i lies inside a cube of side h centered on x and zero otherwise. The total number of data points lying inside this cube will therefore be $K = \sum_{i=1}^{N} k\left(\frac{x-x_i}{h}\right)$. Substituting this expression in Eq. (3) gives the following result for the estimated density at x,

$$p(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h^D}k\left(\frac{x-x_i}{h}\right) \tag{5}$$

where $V = h^D$ for the volume of a hypercube of side h in D dimensions. Eq. (5) represents the kernel density estimator [16]. Even though Gaussian kernels are the most often used, there are various choices among kernels that can be found in the literature [17].

2.2. Domain-based method

Domain-based method requires a boundary to be created based on the structure of the training dataset. These methods are typically insensitive to the specific sampling and the density of the target class because they describe the target class boundary, or the domain, and not the class density. Class membership of unknown data is then determined by their location with respect to the boundary. Domain-based novelty detection is approached with the two-class problem in terms of Support Vector Machine (SVM), where the location of the novelty boundary is determined using only those data that lie closest to it (in a kernel-based transformed space), by means of the support vectors. All other data from the training set (those that are not support vectors) are not considered when setting the novelty boundary. Hence, the distribution of data in the training set is not considered, which is seen as an easy novelty detection approach [7]. The original SVM is a network that is ideally suited for binary pattern classification of data that are linearly separable. Indeed, the SVM defines a hyperplane that maximizes the separating margin between two classes. Since the introduction of the original idea, several modifications and improvements have been made.

2.2.1. Support vector data description

A data domain description method, inspired by the support vector machine approach, called the Support Vector Data Description (SVDD), is used for novelty or outlier detection. The objective is the definition of a spherically shaped decision boundary around a set of measurements by a set of support vectors describing the hypersphere boundary. The method allows the possibility of transforming the data to new feature spaces, where the SVDD can obtain more flexible and more accurate data descriptions. The minimizing problem to delimitate the radius of the hypersphere is expressed as the Lagrangian, $L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j)$, under the constraints of $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i = 1$, where $\alpha_{i,j}$ are the Lagrange multipliers, $x_{i,j}$ are the data training points, the variable C gives the trade-off between simplicity (or volume of the sphere) and the number of errors (number of target objects rejected). For those objects the coefficients $\alpha_{i,j}$ will be nonzero and are called the support objects. In order to determine whether a new measurement is within the hypersphere, the distance to the center of the sphere has to be calculated. A new measurement z is considered *known* when this distance is smaller than the radius,

$$(z \cdot z) - 2 \sum_i \alpha_i (z \cdot x_i) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \leq r^2 \tag{6}$$

where a is the center of the sphere and r is the radius [18]. Kernels could be applied to soften the margins of the sphere, being applied over the measures and data descriptors.

2.2.2. One-class support vector machine

The One-Class SVM, OC-SVM, is based on the definition of the novelty boundary in the feature space corresponding to a kernel, by separating the transformed training data from the origin in the feature space, with the maximum margin. This approach requires fixing a priori

http://dx.doi.org/10.5772/67531

the percentage of positive data allowed to fall outside the description of the *normal* class. This makes the OC-SVM more tolerant to outliers in the *normal* training data. However, setting this parameter strongly influences the performance of this approach. The shape of the domain delimiting the boundaries depends on the kernel selected. Thus, the development of the algorithm is the classic SVM approach. The difference with the other domain-based method approach is that OC-SVM does not consider a specific structure (e.g., a hypersphere) to delimit the domain and therefore does not automatically optimize the model parameters by using artificially generated unlabeled data which are uniformly distributed. The detection of novelty is therefore delimited by,

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) - p \tag{7}$$

where p is an offset. The famous kernel trick is the procedure of using a kernel function in input space, $K(x_i, x)$, to replace the inner product of two vectors into a huge, or even infinite, dimensional feature space. Some drawbacks of these methods are found in literature reviews [7], and it turns out to be surprisingly sensitive to specific choices of representations and kernels in ways which are not very transparent. In addition, the proper choice of a kernel is dependent on the number of features in the binary vector. Since the difference in performance is very dramatic based on these choices, this means that the method is not robust without a deeper understanding of these representation issues.

2.3. Distance-based method

Distance-based methods represent a novelty detection approach similar to that of estimating the PDF of data. Distance-based methods such as nearest neighbors or clustering are based on well-defined distance metrics to compute distance, as the similarity criterion, among data points.

2.3.1. Nearest neighbor

The main idea that rears this technique is that the *normal* data is projected near their neighborhoods, while novelties will be projected far from their neighbors. That is, considering an unknown data point x, this point is accepted as normal if the distance to its nearest neighbor y, in the training set, is less than or equal to the distance from y to the nearest neighbor of y in the training set. Otherwise, x is considered as novelty. Euclidian distance is the most popular choice for univariate and multivariate continuous attributes,

$$\|x - y\| = \sqrt{\sum_{i=1}^{D} (x_i - y_i)^2} \tag{8}$$

Several well-defined distance metrics to compute the distance (or the similarity measure) between two data points can be used, which can broadly be divided into distance-based methods, such as the distance to the kth nearest neighbor and local density-based methods in which the distance to the average of the k's nearest neighbors is considered [11].

In conclusion, novelty detection approaches differ on the assumptions made about the nature of the available data. Each approach exhibits its own advantages and disadvantages and faces different challenges for complex datasets. **Table 1** collects the main characteristics of the considered methods. Thus, probabilistic methods make use of the distribution of the training data to determine the location of the novelty boundary. Domain-based methods determine the location of the novelty boundary using only those data that lie closest to it and do not make any assumptions about the data distribution. Distance-based methods require the definition of an appropriate distance measure for the given data.

Method	Advantages	Disadvantages
Domain-based *i.e. One-class SVM*	Robust to labeled outliers in training by forcing them to lie outside the description. Robust to unlabeled outliers in training.	Several configuration parameters. Sensitive to the scaling of the feature values. Requires a minimum number of training.
Probabilistic parametric *i.e. Gaussian mixture models*	Great advantage when a good probability distribution is assumed. Provides a more flexible density method.	Requires a large number of training samples to overcome the curse of dimensionality. The distribution of the data is assumed. Unlabeled outliers in training affect the estimation of the covariance matrix.
Probabilistic nonparametric *i.e. Kernel density estimator*	Flexible density model. Possible configuration of the kernel width h on each feature direction. Low computational cost for training. The density estimation is only influenced locally.	Requires a large number of training samples to overcome the curse of dimensionality. Expensive computational cost for testing. Limited applicability of the method when there is a large dataset in high-dimensional feature spaces.
Distance-based i.e. k-NN	Rejects parts of the feature space which are within the target distribution. Lack of configuration parameters, besides k; therefore, it relies completely on the training samples.	Scale sensitive due to the use of distances in the evaluation of test objects. Performance affected when unlabeled outliers are presented in training. Sensitive to noise.

Table 1. Summary of the main characteristics of the novelty detection approaches.

3. Case study

In order to illustrate the practical implementation of novelty detection in an industrial application, an interesting case study is proposed next. Indeed, as it has been mentioned, currently, due to the worldwide market situation, the industrial sector is being subjected to a high degree of competitiveness. Critical sectors as the automotive industry are investing in higher levels of quality and safety assessment procedures in order to reduce costs without compromising the attributes of their mechanical manufactured assets. In regard to the automotive rotatory mechanical components, such as the electrical-assisted power steering columns (EPS), end-of-line tests (EOLs) are carried out to analyze their performances. The EPS column is rotated by a test machine in order to quantify the required torque to perform a complete revolution of the EPS column without the influence of any external load. Thus, if the recorded torque is compared with a reference pattern for decision support purposes, then, the EOL test is complete. However, the condition monitoring of the EOL machines, as the represented in **Figure 3**, has

http://dx.doi.org/10.5772/67531

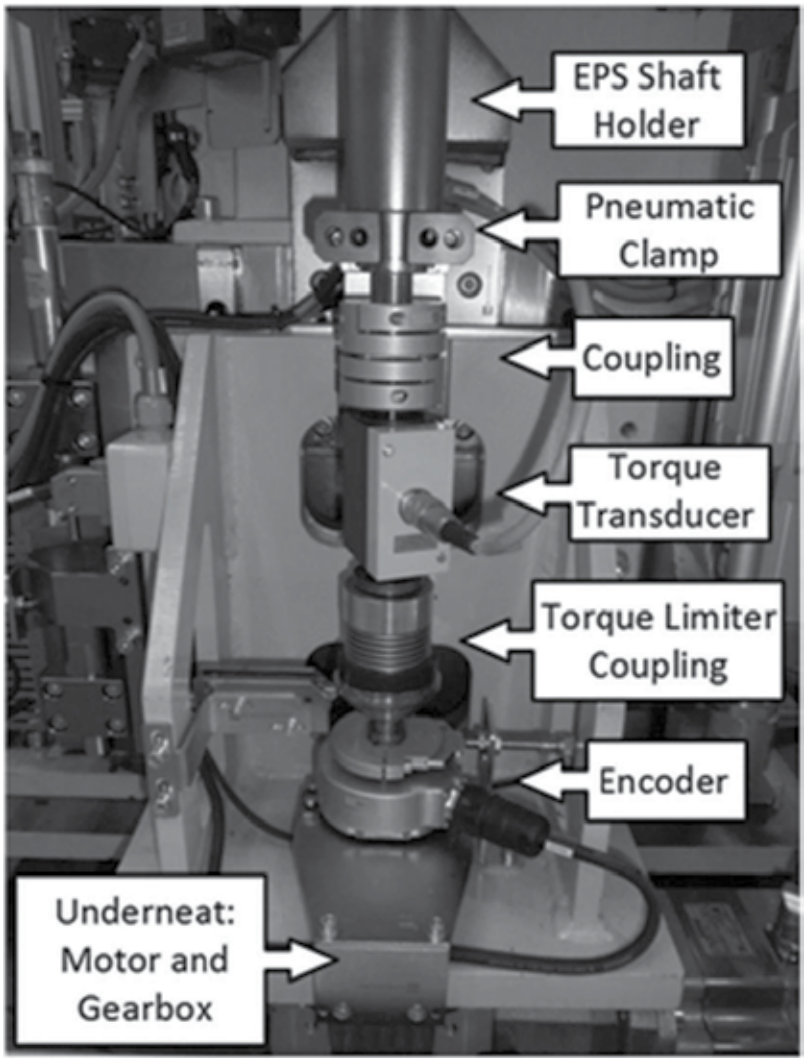


Figure 3. End-of-test machine, composed by a servomotor, a gearbox, an encoder, a torque transducer and a pneumatic clamp to hold the intermediate shaft of the EPS column.

not being attended classically. The maintenance program is limited to a preventive approach, leading to torque response deviations due to EOL machine degradation that are not detected by the machine operator until an evident malfunction. In this regard, the detection and identification of EOL malfunctions during its operation becomes an impactful contribution to the sector and is considered a challenging condition-based monitoring scenario.

In this work, a specific end-of-line test machinery is analyzed. The system under monitoring is based on an electrical drive, where a 1.48 kW at 3000 rpm servomotor connected to a 60:1 reduction gearbox emulates the input torque of the steering wheel to perform a 180° turn in order to evaluate the mechanical performances of power-assisted steering systems. The

measurement equipment is focused on the acquisition of the torque signal and the rotatory shaft position from the encoder. Data acquisition is done at 1 kHz of sampling frequency by means of a NI cDAQ-9188, composed by the modules NI-9411 and NI 9215.

The torque induced by the drive is expected to follow a specific predefined set point pattern. However, these test systems present two main limitations: first, if the test machine does not generate the input of the torque set point correctly, an inaccurate result is obtained during the assessment of the power-assisted steering system, leading to the nonvalidation of the components under test and second, the lack of malfunctions' characterization over the testing machine, since the faults' variability and appearance in the torque generation test are unpredictable. Thus, this work presents an electromechanical system novelty detection approach, based on the temporal torque signal characterization by statistical time features and the evaluation of different novelty detection algorithms (probabilistic, domain-based and distance-based), for novelty assessment.

In order to analyze the performance of the proposed methodology, some faulty conditions have been induced in the machine to provoke different severity degrees of a common fault scenario. Three operating scenarios are considered, that is, healthy, H, a coupling low wear, C_{LW}, and a coupling high wear, C_{HW}. The coupling wear fault is emulated by employing two different intermediate elastomers in the torque limiter coupling, each one with different dynamic torsional stiffness (DTS). The values of the DTS of the pieces under test are all lower than the standard used in the machine in order to emulate classical wear, thus, 2580 Nm/rad corresponds to C_{LW} and 2540 Nm/rad to C_{HW}.

3.1. Method

During the test, the assisting motor of the EPS is not powered. The test starts smoothly in a clockwise direction for the first 45° until a speed set point is reached. The acceleration time depends on the drive capability. During the next 360°, the speed is fixed at the set point, in this case 15 rpm. The last additional 45° is for a mild brake of the EPS column under test. Then, the same procedure is employed to return to the original start point in the opposite direction. The drive is applied to the steering shaft of the EPS. Then, the torque signal analysis is carried out during the stationary speed set point corresponding to a 360° turn of the EPS column. It is expected that malfunctions and anomalies could appear during segments of the revolution of the EPS column; therefore, the segmentation represents a viable strategy to gain resolution during the characterization. That is, the 4-second torque signal (time taken to perform the 360° turn) is segmented in four parts of 1 second. A set of five statistical time-domain features is calculated from each segment of the torque signal. The proposed features are listed in **Table 2**. These features have been successfully employed in different studies for electromechanical systems' fault detection [19]. Therefore, a total of 20 features are calculated from each torque signal measurement.

High-dimensional datasets complicate the learning task of novelty detection as well as multiclass classification methods, because of the possible presence of nonsignificant and redundant information in the data, compromising the proper convergence of the algorithms. Indeed, the empty space phenomenon states that to cover the whole space, it needs a number of samples that grows

http://dx.doi.org/10.5772/67531

Root mean square (RMS)	$\sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k)^2}$	Variance	σ^2
Shape factor	$\frac{RMS}{\frac{1}{n}\sum_{k=1}^{n}\lvert x_k\rvert}$	Skewness	$\sum_{k=1}^{n}\frac{(x_k-\bar{x})^3}{n\sigma^3}$
Crest factor (CF)	$\frac{\max(x)}{RMS}$	Kurtosis	$\sum_{k=1}^{n}\frac{(x_k-\bar{x})^4}{n\sigma^4}$

Table 2. Statistical time-domain features used for torque signal characterization.

exponentially with dimensionality. Thus, the curse of dimensionality implies that in order to learn successfully, it needs a number of training examples that also grows exponentially with the dimensionality. The "concentration of measure" phenomenon seems to render distance measures not relevant to whatever concept is to be learned as the dimension of the data increased. For these reasons, there is a necessity to apply dimensionality reduction techniques in condition monitoring applications. Thus, in order to analyze the performance of different novelty detection approaches, two main dimensionality reduction approaches are applied over the 20-dimensional vectors, that is, Principal Component Analysis, PCA, and Laplacian Score, LS.

Indeed, the dimensionality reduction strategies differ in the criteria applied over the data in order to reach a reduced feature space. PCA is one of the most commonly used techniques for unsupervised dimensionality reduction. It aims to find the linear projections that best capture the variability of the data [13]. Another well-known technique is the LS, where the merit of each feature is measured according to its locality preservation power. A nearest neighbor-based graph is constructed from the training set and analyzed to rank each feature individually according to a weighting approach selected for the graph's edges. To rank each feature, its LS is computed, which is a measure of the extent to which the analyzed feature preserves the structure present in the graph divided by the variance of the feature. For a feature to be selected, it must have a low LS, which implies high variance and locality [20].

Finally, the necessity of evaluating the novelty detection performance is critical. The use of a particular score depends on multiple interests, and then, the analysis of complementary scores represents the most interesting solution. Next, the most useful and common scores in a discrete scenario are described in order to be used later during the analysis of the experimental results.

- Accuracy and classification error (1-accuracy): One of the most frequent scores used to evaluate discrete classification in electromechanical diagnosis is accuracy. This score is indicative of the classification error committed while evaluating, in our case, two classes,

$$Accuracy = \frac{FP + FN}{N} \tag{9}$$

where FP is the number of false positives, FN is the number of false negatives and N, the total number of analyzed measures. Two novelty detection approaches could exhibit the same accuracy but provide a different novelty ratio for each class (normal data and novelty data).

- True positive rate (recall or sensitivity): This measure provides a proportion of one kind of sample that was correctly assessed. But it only evaluates the positive cases,

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

where TP is the number of true positives.

- Precision: This performance metric evaluates the correct classification of the positive class,

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

- F-measure: This score can help to solve any contradiction that may appear between precision and recall scores. F-measure leaves out the TN performance. Several versions exist. The most common expression is,

$$F_1 = \frac{2[Precision * Recall]}{[2 * Precision + Recall]} \tag{12}$$

3.2. Experimental results

In order to expose the novelty detection performances, the outline of the experimental results is presented as follows: The initial database is characterized by the proposed set of features, then both feature reduction approaches are applied and, finally, over each reduced set of features, the three novelty detection approaches are applied. The application of the novelty detection is done sequentially, that is, first, the data corresponding to the healthy, H, operating scenario is characterized by the novelty model. Second, the first fault operating scenario, C_{LW}, is presented as well as additional measures of the H operating scenarios. At this point, the performance of the novelty detection model is analyzed. Third, the novel data identified is included in the upgraded version of the novelty model by retraining, and over this updated novelty model, the second fault operating scenario, C_{HW}, is presented as well as the additional measures of the C_{LW} and H operating scenarios. At this point, the performance of the novelty detection models is analyzed again. Finally, the novel data identified is included in an upgraded version of the novelty model by a new retraining.

Three novelty detection methods have been implemented, that is, the mixture of Gaussians as the probabilistic approach, one-class support vector machines as the domain-based novelty detection approach and, finally, k-nearest neighbors as the distance-based novelty detection approach. Next, the PCA variant of the novelty detection methodology is shown in **Figure 4**.

The proposed scores during the assessment of the novelty detection models in front of a new set of measurements are shown next. Thus, in **Table 3**, the scores in regard to the PCA feature reduction and the novelty detection models 'performance, dealing with the projection of new measurements corresponding to H and C_{LW} operating scenarios over the novelty models

http://dx.doi.org/10.5772/67531

trained with the H operating scenario, can be seen. It should be noted that, in regard to all the scores shown in the next tables, a 10-fold cross validation strategy has been considered, in which the mean and the dispersion ratio of the obtained scores is shown.

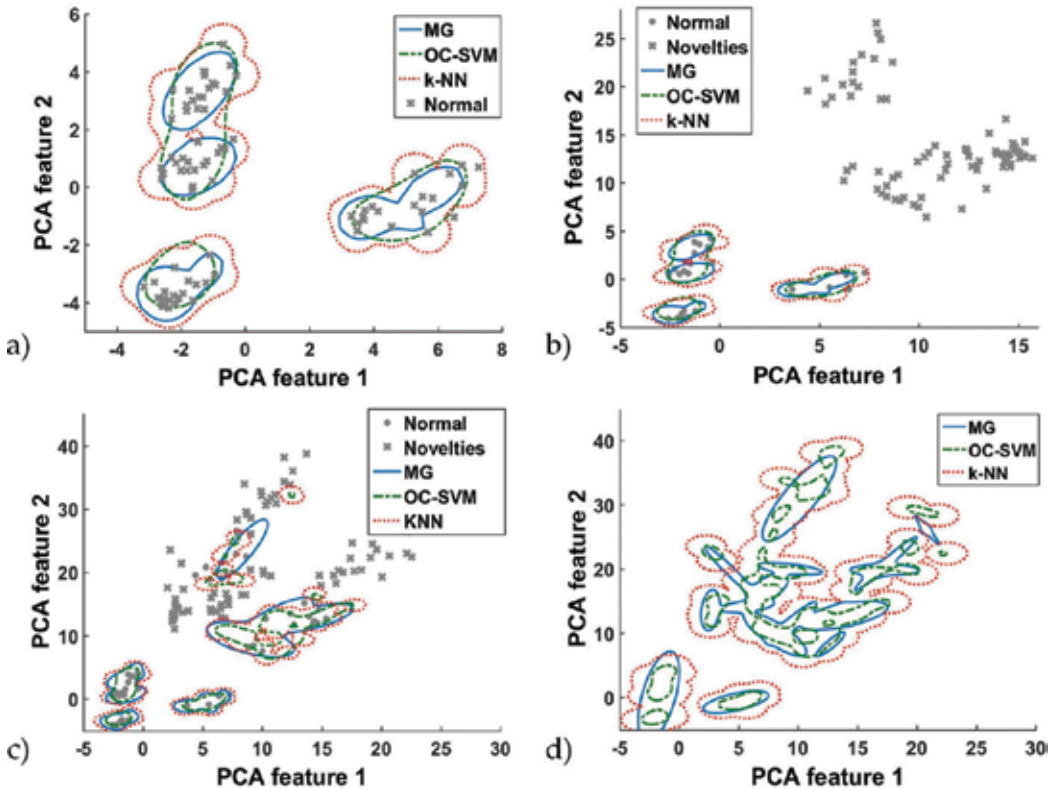


Figure 4. Performance of the three novelty detection models in a PCA-based 2-dimensional feature space. MG mixture of Gaussian, OC-SVM, one-class support vector machine and *k*-NN, *k*-nearest neighbor. (a) Novelty models' boundaries during the characterization of the H operating scenario. (b) Projection of new measurements corresponding to H and C_{LW} operating scenarios over the novelty models trained with the H operating scenario. (c) Novelty model boundaries' update by the incorporation of the H and C_{LW} operating scenarios' measures detected as novelties and the projection of new measurements corresponding to H, C_{LW} and C_{HW} operating scenarios over the novelty models. (d) Novelty model boundaries' update by the incorporation of the H, C_{LW} and C_{HW} operating scenarios' measures detected as novelties.

	MG	**OC-SVM**	**k-NN**
Accuracy	0.912 (±0.033)	0.952 (±0.025)	0.950 (±0.034)
Recall	1.000 (±0.000)	1.000 (±0.000)	1.000 (±0.000)
Precision	0.898 (±0.035)	0.942 (±0.029)	0.949 (±0.038)
F1 score	0.946 (±0.019)	0.970 (±0.015)	0.974 (±0.02)

Table 3. Associated scores to the PCA and novelty detection performance dealing with the class H as normal and C_{LW} as novelty.

Also, in **Table 4**, the scores in regard to the PCA feature reduction and the novelty detection models' performance, dealing with the projection of new measurements corresponding to H, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the H and C_{LW} operating scenarios, can be seen.

	MG	**OC-SVM**	**k-NN**
Accuracy	0.930 (±0.025)	0.863 (±0.014)	0.966 (±0.014)
Recall	0.997 (±0.005)	0.995 (±0.006)	0.985 (±0.005)
Precision	0.902 (±0.033)	0.822 (±0.013)	0.962 (±0.022)
F1 score	0.947 (±0.018)	0.901 (±0.01)	0.973 (±0.010)

Table 4. Associated scores to the PCA and novelty detection performance dealing with the H and C_{LW} as known operating scenarios and C_{HW} as novelty.

Similarly, next, the LS variant of the novelty detection methodology is shown in **Figure 5**.

The proposed scores during the assessment of this variant of novelty detection models in front of a new set of measurements are also shown next. Thus, in **Table 5**, the scores in regard to the LS feature reduction and novelty detection models' performance, dealing with the projection of new measurements corresponding to H and C_{LW} operating scenarios over the novelty models trained with the H operating scenario can be seen.

	MG	**OC-SVM**	**k-NN**
Accuracy	0.956 (±0.021)	0.971 (±0.026)	0.967 (±0.028)
Recall	1.000 (±0.000)	1.000 (±0.000)	1.000 (±0.000)
Precision	0.940 (±0.023)	0.965 (±0.031)	0.960 (±0.033)
F1 score	0.969 (±0.012)	0.982 (±0.016)	0.980 (±0.017)

Table 5. Associated scores to the LS and novelty detection performance dealing with the class H as normal and C_{LW} as novelty.

Also, in **Table 6**, the scores in regard to the LS feature reduction and novelty detection models' performance, dealing with the projection of new measurements corresponding to H, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the H and C_{LW} operating scenarios can be seen.

	MG	**OC-SVM**	**k-NN**
Accuracy	0.910 (±0.016)	0.841 (±0.032)	0.890 (±0.014)
Recall	0.955 (±0.017)	0.925 (±0.021)	0.865 (±0.044)
Precision	0.922 (±0.027)	0.840 (±0.044)	0.940 (±0.030)
F1 score	0.938 (±0.012)	0.879 (±0.021)	0.899 (±0.014)

Table 6. Associated scores to the LS and novelty detection performance dealing with the H and C_{LW} as known operating scenarios and C_{HW} as novelty.

http://dx.doi.org/10.5772/67531

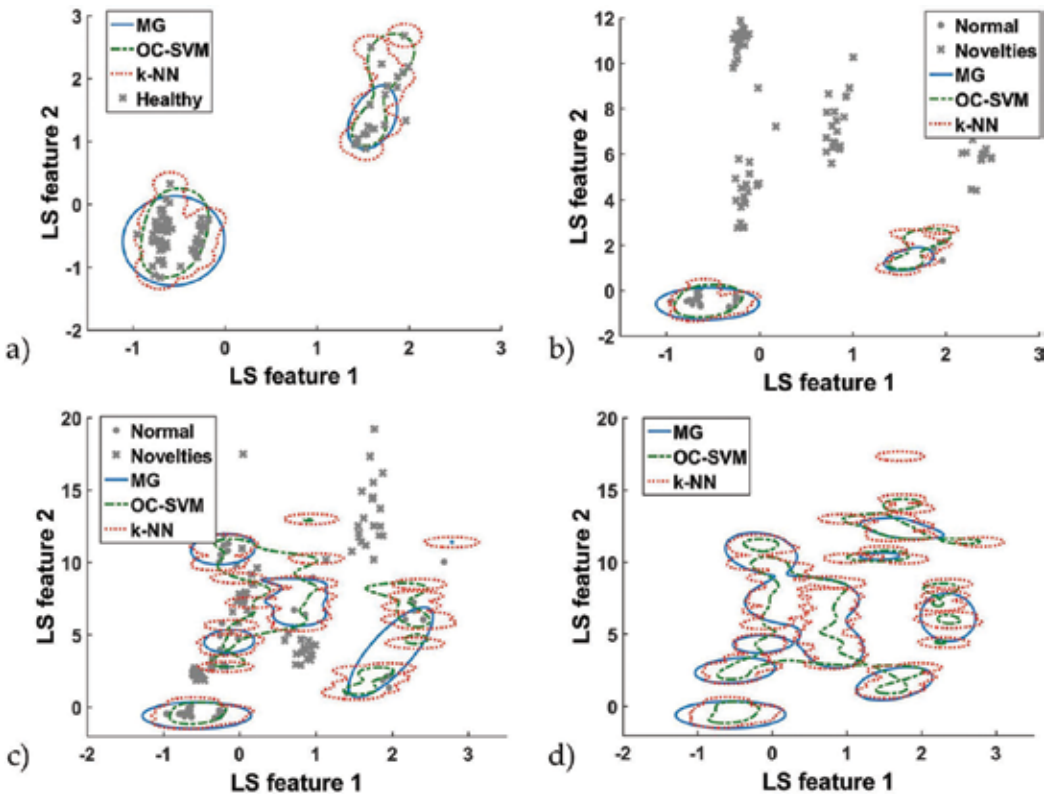


Figure 5. Performance of the three novelty detection models in a LS based 2-dimensional feature space. MG mixture of Gaussian, OC-SVM, one class support vector machine, and *k*-NN, *k*-nearest neighbor. (a) Novelty models boundaries during the characterization of the *H* operating scenario. (b) Projection of new measurements corresponding to *H* and C_{LW} operating scenarios over the novelty model trained with the *H* operating scenario. (c) Novelty model boundaries update by the incorporation of the *H* and C_{LW} operating scenarios measures detected as novelties, and projection of new measurements corresponding to *H*, C_{LW} and C_{HW} operating scenarios over the novelty models. (d) Novelty model boundaries update by the incorporation of the *H*, C_{LW} and C_{HW} operating scenarios measures detected as novelties.

In regard to the feature reduction effect over the novelty detection performance, it has been taken into account that both methods, PCA and LS, represent linear approaches to the reduction of the initial 20-dimensional feature set. This premise is not a limitation to the analysis of the novelty detection models considered; however, the feature space could be further improved in order to maximize the obtained results for the specific application.

Independently of the novelty detection model, the first test stage, that is, the assessment of new measurements corresponding to *H* and C_{LW} operating scenarios over the novelty models trained with the *H* operating scenario, shows a clear superiority of the LS approach. The accuracy obtained with the LS approach reaches till 97% and is, in all cases, better than using PCA that reaches a maximum of 92%. However, the second test stage, that is, the assessment of new measurements corresponding to *H*, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the *H* and C_{LW} operating scenarios, shows a clear superiority of the PCA approach. The accuracy obtained with the PCA approach reaches till 96% and is, in most of the cases, better than using LS that reaches a maximum of 91%.

This effect is reasonable in dealing with the available data because the LS approach allows a better representation in terms of novelty detection. That is, considering that all the available data corresponds to the unique class *normal* or *known*, the performance of the novelty detection model will be enhanced if less data dispersion is presented. In this sense, the LS feature space shows a more compact projection of the data, at least during the first test stage, a fact that facilitates the definition of the novelty detection boundaries and the posterior accuracy. However, dealing with the second test stage, the maximization of the variance by means of the PCA avoids false negatives. In fact, the dispersion of data is desired when complexity of data is considered, since new operating scenarios could be assessed as known measurements. This performance of the feature reduction techniques over the novelty detection performance is a critical aspect during the condition monitoring configuration, since a trade-off between feature space complexity and data dispersion must be reached. Nevertheless, in this case study, the proposed novelty detection methodology including both the feature reduction techniques exhibits high ratios of performances.

In regard to the novelty detection models, independently of the feature reduction technique, the first test stage, that is, the assessment of new measurements corresponding to H and C_{LW} operating scenarios over the novelty models trained with the H operating scenario, shows a clear superiority of the OC-SVM and k-NN approaches in terms of accuracy, precision and F1-score, considering that the recall is maximum in all three cases. However, the second test stage, that is, the assessment of new measurements corresponding to H, C_{LW} and C_{HW} operating scenarios over the novelty models trained with the H and C_{LW} operating scenarios, shows a superiority, of the k-NN approach, mainly in terms of accuracy and precision, although the MG shows also good behavior in terms of recall.

In fact, as it has been mentioned, the probabilistic novelty detection approach, represented by the MG technique, assumes a data dispersion that, dealing with unknown operating scenarios, cannot be the optimum. This fact is smoothened when the data density increases, since more information is available in order to infer a proper PDF. In case of OC-SVM and k-NN, both techniques show wide novelty detection boundaries, which allow a better characterization of the data distribution by means of good generalizations. However, it must be taken into account that, qualitatively, a more complex partition of the feature space is reached by the k-NN, and although, as it has been explained, this can be controlled by the value of k, such tuning is not trivial and, then, OC-SVM represents a more simple solution.

4. Conclusions

A condition monitoring scheme for novelty detection is applied to an industrial end-of-line test machinery of electrical-assisted power steering columns, where the *healthy* data is the initial available information. The fault conditions considered consist of two severities of one commonly presented fault in the mechanical parts of the electrical drive of the test machine, coupling wear. The fault condition is presented in two stages, in order to analyze the detection and learning capabilities of the considered approaches. These fault severities represent a challenge for the data analysis due to the similitudes between the torque signals characterizing each fault.

http://dx.doi.org/10.5772/67531

Six variants of the methodology are proposed and analyzed. Thus, two feature reduction approaches by means of PCA and LS are considered in order to emphasize the information contained in the 20-dimensional vectors of statistical time-based features in which each torque measurement is characterized. Later, three novelty detection modelling approaches are introduced and implemented, that is, the probabilistic method by means of the mixture of Gaussians, domain-based methods by means of one-class support vector machine and distance-based methods, by means of *k*-nearest neighbors. A comparison and analysis between the novelty models and the feature reduction procedures is performed to analyze the proper selection of novelty models for these scenarios. The results have shown that the combination of PCA as feature reduction and *k*-NN as the novelty detection model reaches, in general, the best-considered scores, mainly the accuracy, 96 and 90%, and precision, 96 and 94%. However, the OC-SVM alternative must also be considered due to its simpler configuration requirements and good performances.

Acknowledgements

The authors gratefully acknowledge the financial support from the MINECO of Spain, under the Project CICYT TRA2013- 46757-R, the Generalitat de Catalunya under the Project GRC MCIA, Grant n° SGR 2014-101, and the CONACyT of Mexico under the scholarship 313604 grant. Also, the authors would like to thank the support and the access to the friction test machine database provided by MAPRO Sistemas de ensayo S.A., especially to Álvaro Istúriz and Alberto Saéz.

Author details

Miguel Delgado Prieto[1]*, Jesús A. Cariño Corrales[1], Daniel Zurita Millán[1],
Marta Millán Gonzalvez[2], Juan A. Ortega Redondo[1] and René de J. Romero Troncoso[3]

*Address all correspondence to: miguel.delgado@mcia.upc.edu

1 MCIA Research Center, Technical University of Catalonia, Terrassa, Spain

2 MAPRO Test and Assembly Systems S.A., Sant Fruitós de Bages, Spain

3 University of Guanajuato, Salamanca, Mexico

References

[1] Yin S., Ding S. X., Xie X., Luo. H. A review on basic data-driven approaches for industrial process monitoring. IEEE Transactions on Industrial Electronics. 201;**61**(11):6418–6428. doi:10.1109/TIE.2014.2301773

[2] Davies, A., editor. Handbook of condition monitoring: Techniques and methodology. 1st ed. Ireland: Springer Science & Business Media Dordrecht; 1998. 564 p. doi:10.1007/978-94-011-4924-2

[3] Ahmad R., Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. Computers & Industrial Engineering. 2012;**63**(1):135–149. doi:10.1016/j.cie.2012.02.002

[4] Boashash B., editor. Time-frequency signal analysis and processing: A comprehensive reference. 2nd ed. UK: Academic Press; 2016. 1020 p

[5] Möller D.P.F. Digital manufacturing/industry 4.0. In: Möller D.P.F., editor. Guide to computing fundamentals in cyber-physical systems. 1st ed. Switzerland: Springer; 2016. pp. 307–375. doi:10.1007/978-3-319-25178-3_7

[6] Teti R., Ferretti S., Caputo D., Penza M., D'Addona D. M. Monitoring systems for zero defect manufacturing. In: Teti R., editor. Eighth CIRP conference on intelligent computation in manufacturing engineering, 18 July, Naples. Procedia CIRP; 2013. pp. 258–263. doi:10.1016/j.procir.2013.09.045

[7] Pimentel M. A. F., Clifton D. A., Clifton L., Tarassenko L. A review of novelty detection. Signal Processing. 2014;**99**(6):1165–1684. doi:10.1016/j.sigpro.2013.12.026

[8] Jumutc V., Suykens J. A. K. Multi-class supervised novelty detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014;**36**(12):2510–2523. doi:10.1109/TPAMI.2014.2327984

[9] Jyothsna V., Rama Prasad V. V. A review of anomaly based intrusion detection systems. International Journal of Computer Applications. 2011;**28**(7):26–35

[10] Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. ACM Computing Surveys. 2009;**41**(3):15–58. doi:10.1145/1541880.1541882

[11] Markou M., Singh S. Novelty detection: A review—Part 1: Statistical approaches. Signal Processing. 2003;**83**(12):2481–2497

[12] Markou M., Singh S. Novelty detection: A review—Part 2: Neural Network Based Approaches. Signal Processing. 2003;**83**(12):2499–2521. doi:10.1016/j.sigpro.2003.07.019

[13] Duda R. O., Hart P. E., Stork D. G., editors. Pattern classification. 2nd ed. Chichester: Wiley; 2000. 680 p

[14] Reynolds D. Gaussian mixture models. In: Li S. Z., Jain A. K., editors. Encyclopedia of biometrics. 2nd ed. United States: Springer US; 2015. pp. 827–832. doi:10.1007/978-1-4899-7488-4_196

[15] Parra L., Deco, G., Miesbach, S. Statistical independence and novelty detection with information preserving nonlinear maps. Neural Computation. 1996;**8**(2):260–269. doi:10.1162/neco.1996.8.2.260

http://dx.doi.org/10.5772/67531

[16] Elmoataz A., Lezoray O., Nouboud F., Mammass D., Meunier J., editors. Image and signal processing. 1st ed. New York: Springer; 2010. 602 p

[17] Bishop C. M., editor. Pattern recognition and machine learning. 1st ed. New York: Springer-Verlag; 2006. 740 p

[18] Tax D. M. J., Duin R. P. W. Support vector domain description. Pattern Recognition Letters. 1999;**20**(11–13):1191–1199. doi:10.1016/S0167-8655(99)00087-2

[19] Delgado-Prieto M., Cirrincione G., Espinosa A. G., Ortega J. A., Henao H. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. IEEE Transactions on Industrial Electronics. 2013;**60**(8):3398–3407. doi:10.1109/TIE.2012.2219838

[20] He X., Cai D., Niyogi P. Laplacian score for feature selection. In: Weiss Y., Schölkopf B., Platt J. C., editors. Neural information processing systems. Vancouver: MIT Press; 2005. pp. 1–8

Chapter 6

Fault Diagnosis and Health Assessment for Rotating Machinery Based on Kernel Density Estimation and Kullback-Leibler Divergence

Yu Liu, Chen-Yao Yan and Fan Zhang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67360

Abstract

To avoid severe damages and unexpected shutdowns, fault diagnosis and health assessment of rotating machinery have received considerable attention in recent years. On the other hand, as a great amount of data become acquirable and accessible in industry, data-driven tools have become an emerging research area, acting as a complement to the model-based (or physics-based) fault diagnosis and health assessment methods. In this chapter, based on the kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID), a new data-driven fault diagnosis approach and a new health assessment approach are introduced. By utilizing the KDE, the statistical distribution of selected features can be readily estimated without assuming any parametric family of distributions, whereas the KLID is able to quantify the discrepancy between two probability distributions of selected features. An integrated Kullback-Leibler divergence, which aggregates the KLID of all the selected features, is introduced to discriminate various fault types or health status of rotating machinery. The effectiveness of the proposed approaches is demonstrated through three case studies of fault diagnosis and health assessment of rotating machinery.

Keywords: data-driven approach, fault diagnosis, health assessment, kernel density estimation, kullback-Leibler divergence, rotating machinery

1. Introduction

Rotating machinery has widespread applications in advanced manufacturing and engineering systems, e.g., wind turbines, power generators, and machining tools. The crucial components in rotating machinery, such as bearings and gears, are oftentimes suffering undesirable

stresses and sudden shocks under which initial defects will appear [1]. If maintenance activities cannot be taken properly and timely, tiny defects will gradually propagate and eventually cause severe damages and unexpected shutdown to the entire systems. It is, therefore, of paramount importance to accurately detect the presence of faults as early as possible and track the growth of the tiny faults to avoid the consequence of severe damages caused by faults and also facilitate preventive maintenance planning before the complete failure of engineering systems [1].

Fault diagnosis and health assessment are the two important tools for detecting the operating condition of rotating machinery based on which preventative maintenance can be scheduled in a timely manner. In general, existing methods for fault diagnosis and health assessment can be classified into two categories [2]: model-based (or physics-based) approaches and data-driven approaches. The model-based approaches require specific mechanistic knowledge and theory relevant to the monitored machine, and a particular fault or health status of a system can be determined by comparing available system measurements with a priori information represented by the corresponding system's analytical/mathematical model [3]. These methods could be very accurate when a correct model can be built up. For example, several models have been developed to characterize the crack growth [4, 5]. However, due to the limited knowledge of underlying mechanisms and physics, it becomes very difficult, or even impossible, to fully understand the evolution of defects and faults of complex engineered systems [1, 2]. With the fast development of condition monitoring and intelligent computing technologies, data-driven approaches have received considerable attention in recent years. Many advanced classification methods have been applied to data-driven fault diagnosis [6–9]. Among them, support vector machine [9, 10] and artificial neural network (ANN) [11] are two representative and powerful classification methods, and they have been extensively used in fault diagnosis for rotating machinery [9–13]. By using the data-driven approaches, the fault type or health status of a system can be mapped into the feature space [1, 2]. In other words, the relation between features extracted from condition monitoring data and fault model/damage levels can be acquired from a set of training data. Thereby, compared to the model-based approaches, the data-driven approaches possess two merits: (1) by the data-driven approaches, fault diagnosis or health assessment can be executed automatically without heavy involvement of engineers and (2) unlike the model-based approaches that need professional expertise to make judgments, the data-driven approaches do not rely on expertise and knowledge from experts too much [13].

In most cases, a data-driven approach for fault diagnosis and health assessment of rotating machinery consists of five basic steps as shown in **Figure 1**. The raw data, e.g., vibration signals, collected from condition monitoring serve as inputs of a data-driven fault diagnosis or health assessment approach. Subsequently, by using advanced signal processing algorithms, e.g., the fast Fourier transform (FFT), the empirical model decomposition (EMD), and the wavelet transform, a bunch of features which are more or less relevant to the health status of the monitored device can be extracted from the raw data. A subset of the most significant features, which are sensitive to a specific fault type or health status of the system, will be chosen from all the extracted features. On the other hand, irrelevant and redundant features can be eliminated at this stage to mitigate the computational burden and improve the accuracy

http://dx.doi.org/10.5772/67360

of results. It is followed by the fault classification or health assessment where the selected features will be used as inputs of fault/health status classifier. Many advanced classification methods can be applied. Among many, the support vector machine (SVM) [10] and the artificial neural network (ANN) [11] are two representative approaches.

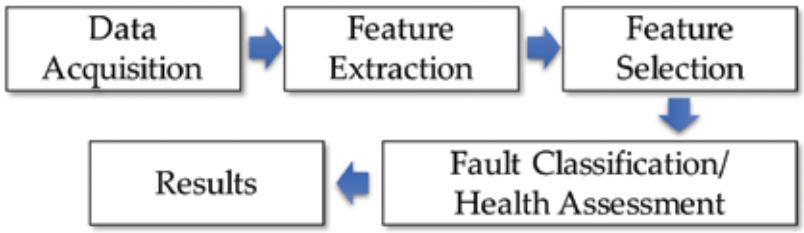


Figure 1. The basic procedures of data-driven fault diagnosis or health assessment methods.

In this chapter, a new data-driven fault diagnosis approach and a data-driven health assessment approach are put forth. Two statistical tools, i.e., the kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID), are used jointly to identify fault modes/health status of rotating machinery from a statistical viewpoint. The KDE, which is a nonparametric probability density estimation approach, is able to adaptively fit a data set to a smooth density function without pre-specifying a particular distribution type [14, 15]. On the other hand, the KLID, which is so-called information divergence or relative entropy, is a measure of the discrepancy between two probability distributions [16]. By using the KDE and KLID jointly, an integrated Kullback-Leibler divergence can be developed to identify faults modes/health status of rotating machinery.

The rest of this chapter is rolled out as follows: Section 2 introduces the principle of the KDE and the KLID. The proposed fault diagnosis approach together with two case studies is presented in Section 3. In Section 4, the proposed health assessment approach and its application to a case study are elaborated. The chapter closes with a brief conclusion in Section 5.

2. The kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID)

2.1. The kernel density estimation

The kernel density estimation originally introduced by Rosenblatt and Parzen [14, 15] is a nonparametric tool to infer the probability density function of a data set. It stems from the empirical probability density function.

Let X_1, X_2, ⋯, X_n represent n independent and identically distributed (i.i.d.) random samples from a random quantity X with an unknown probability density function $f(x)$. The kernel density function is defined as:

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \tag{1}$$

where $K(\cdot)$, a symmetric function with integration equal to 1, is the kernel function. The kernel function may not be necessary a position function but has to guarantee $\widehat{f}_h(x)$ satisfies the basic requirement of a probability density function. Many different types of kernel functions have been proposed in Ref. [15], e.g., uniform, Gaussian, triangle, Epanechnikov, and quaritic. Particularly, the Gaussian kernel function, which has been extensively adopted due to many mathematical properties, such as centrality and gradual decay, is formulated as:

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right) \tag{2}$$

The bandwidth h ($h > 0$) of the kernel function has a heavy influence on the smoothness of $\widehat{f}_h(x)$. A larger h indicates that a greater region of samples around the centre point x influences the probability density estimation, vice versa. A proper setting for the bandwidth h is, therefore, of great significance for the KDE. The mean integrated squared error (MISE) is the most common optimality criterion to choose a proper bandwidth, and it is defined as [15]:

$$MISE(h) = E\int \left(\widehat{f}_h(x) - f(x)\right)^2 dx \tag{3}$$

Under weak assumptions on $f(\cdot)$ and $K(\cdot)$, one has:

$$MISE(h) = AMISE(h) + o\left(\frac{1}{nh} + h^4\right) \tag{4}$$

where $o(\cdot)$ is infinitesimal. The AMISE is the asymptotic MISE, and it is defined as:

$$AMISE(h) = \frac{R(K(\cdot))}{nh} + \frac{1}{4}m_2(K(\cdot))^2 h^4 R\left(f''(\cdot)\right) \tag{5}$$

where $R(g) = \int g(x)^2 dx$; $m_2(K) = \int x^2 K(x)dx$; $f''(\cdot)$ is the second-order derivative of $f(\cdot)$; and n is the total number of samples. The following differential equation can be used to seek the minimal value of the AMISE as:

$$\frac{\partial}{\partial h}AMISE(h) = -\frac{R(K(\cdot))}{nh^2} + m_2(K(\cdot))^2 h^3 R\left(f''(\cdot)\right) = 0 \tag{6}$$

Thus, the minimal value of h is:

$$h^*_{AMISE} = \frac{R(K(\cdot))^{1/5}}{m_2(K(\cdot))^{2/5} R\left(f''(\cdot)\right)^{1/5} n^{1/5}} \tag{7}$$

It should be noted that the above equation is implicit and contains the unknown density function $f(\cdot)$ or $f''(\cdot)$. In engineering practice, the density to be estimated is also Gaussian if the Gaussian basis function is used to approximate univariate data. Thus, the optimal value of h is:

http://dx.doi.org/10.5772/67360

$$h^*_{AMISE} = \left(\frac{4\widehat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\widehat{\sigma}n^{-\frac{1}{5}} \tag{8}$$

where $\widehat{\sigma}$ is the standard deviation of samples. Such approximation called the Gaussian approximation is adopted in this work.

2.2. The Kullback-Leibler divergence

The Kullback-Leibler divergence (KLID) was first introduced by Solomon Kullback and Richard Leibler in 1951 [16], and it has been applied to quantify the difference of two distributions. For two discrete probability distributions P and Q, the KLID of Q from P is written as:

$$D_{KL}(P||Q) = \sum_i \ln\left(\frac{P(i)}{Q(i)}\right)P(i) \tag{9}$$

In essence, Eq. (9) is the expectation of the logarithmic difference between the probabilities P and Q, and the expectation is taken by the probability P. The KLID is valid if the integration of P and Q is both equal to 1. If $Q(i) = 0$, then $P(i) = 0$ for all i. For the case where $P(i) = 0$ and $P(i)/Q(i) = 0$, $\ln(P(i)/Q(i))P(i) = 0$ since $\lim_{x\to 0} x\ln(x) = 0$.

Based on the Gibbs' inequality, $D_{KL}(P||Q) = 0$ if and only if $P = Q$ holds almost everywhere. A smaller value of $D_{KL}(P||Q)$ represents a greater similarity between the two probability distributions. It is noteworthy that although the KLID can quantify the distance between two probability distributions, it does not fully satisfy some important properties of distance measure, e.g., symmetry and triangle inequality. For instance, the KLID of P over Q is generally not exactly equal to the KLID of Q over P. Nevertheless, the symmetry property is very crucial in the classification issue. In our work, the symmetrized distance of KLID defined in Ref. [12] is adopted to measure the discrepancy between two probability distributions, and it is formulated as:

$$D_{KL}(P, Q) = \frac{1}{2}[D_{KL}(P||Q) + D_{KL}(Q||P)]. \tag{10}$$

3. The proposed fault diagnosis approach

Follow the basic procedures of data-driven fault diagnosis method as shown in **Figure 1**, the proposed approach for the fault diagnosis of rotating machinery is given in **Figure 2**. A set of time- and frequency-domain features will be first extracted from the raw vibration signals by the ensemble empirical mode decomposition (EEMD), the Hilbert Transform, and so on. The distance-based feature selection method will be used to identify a subset of sensitive features. The kernel density estimation (KDE) and the Kullback-Leibler divergence (KLID) introduced in Section 2 will be used together as a new classifier to discriminate various fault types.

Figure 2. The procedures of the proposed fault diagnosis approach for rotating machinery.

3.1. The details of the proposed method

In this section, feature extraction, feature selection, kernel density estimation, and Kullback-Leibler divergence will be integrated together to realize the fault diagnosis for rotating machinery. Some important symbols to be used hereinafter are explained here:

1. KD_i^j (j = 1, 2, ⋯, n; i = 1, 2, ⋯, C) denotes the KDE function of the jth feature of the training samples for type i fault. The vector $\mathbf{KD}_i = \left(KD_i^1, KD_i^2, \cdots, KD_i^n\right)$ is the KDE function set of all the n selected features of the training samples for type i fault.

2. TKD_i^j(j = 1, 2, ⋯, n;i = 1, 2, ⋯, C) is the KDE function of the jth feature of the training samples for type i fault after adding a testing sample. The vector $\mathbf{TKD}_i = \left(TKD_i^1, TKD_i^2, \cdots, TKD_i^n\right)$ is the KDE function set of all the n selected features of the training samples for type i fault after a testing sample is added.

3. KL_i^j (j = 1, 2, ⋯, n; i = 1, 2, ⋯, C) is the KLID between KD_i^j and TKD_i^j. The vector $\mathbf{KL}_i = \left(KL_i^1, KL_i^2, \cdots, KL_i^n\right)$ contains the KLIDs of all the n selected features.

The overall flowchart of the proposed approach for classifying two fault types is shown in **Figure 3**.

In **Figure 3**, the proposed method is illustrated through classifying two fault models, i.e., type I fault and type II fault. The sample sets from these two types of fault modes act as the training sample sets, whereas one sample set with unknown fault mode serves as the testing sample set to be classified. Nine time-domain features together with 10 frequency-domain features are extracted from the raw vibration signal and the first four IMFs decomposed by the EEMD. The technical details of the EEMD can be found in Refs. [17, 18]. Thus, the original feature set consists of 95 features. The distance-based evaluation approach is, then, applied to assess the effectiveness of each feature. The corresponding effectiveness factor of the jth (j = 1, 2, ⋯, 95) feature is denoted as α_j (see Ref. [19] for more details on the distance-based evaluation approach). The features with a larger effectiveness factor are more sensitive to these particular fault types. By sorting all the features by their effectiveness factors in a descending order, the first m features are selected from the original feature set and serve as the inputs of the classifier. Thereby, the importance of the ^{j}th feature to the fault classification is formulated as:

http://dx.doi.org/10.5772/67360

$$F_j = \frac{\alpha_j}{\sum_{i=1}^{n} \alpha_i}, (j = 1, 2, \cdots, m) \tag{11}$$

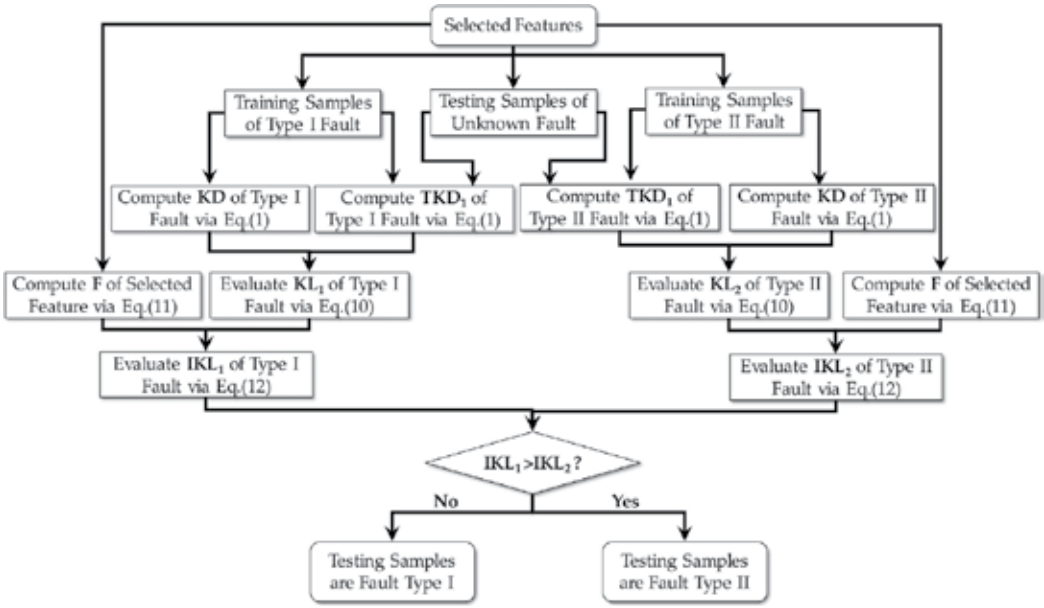


Figure 3. The flowchart of the proposed fault diagnosis approach for the case with two fault types.

The probability density of the selected features of each training set can be then characterized by the kernel density function. For instance, KD_1^1 and KD_2^1 are the first feature of type I and type II faults, respectively, are shown in **Figure 3**. If one sample from the testing sample set is added into the two training sets, respectively, and the corresponding probability distributions of the two new sample sets of the first feature can be also estimated by the kernel density function and denoted as TKD_1^1 and TKD_2^1, respectively. In the same manner, KD_1^j, KD_2^j, TKD_1^j, and TKD_2^j (j = 1,2,3,....,m) for all the selected features can be estimated. It is followed by computing the KL_1^j and KL_2^j, the symmetrized Kullback-Leibler divergences (KLIDs) between KD_1^j and TKD_1^j, KD_2^j and TKD_2^j (j = 1, 2, ⋯, m), via Eq. (10). To get an overall assessment of all the n selected features, an integrated KLID, denoted as IKL_i, is defined here to aggregate all the symmetrized KLIDs KL_i^j for the type i fault together as:

$$IKL_i = \sum_{j=1}^{m} F_i \times KL_i^j \tag{12}$$

where F_j(j = 1, 2, ⋯, m) computed by Eq. (11) is the importance of the jth feature and **F** = (F_1, F_2, ⋯, F_m). Using Eq. (12), IKL_1 and IKL_2 of any sample from the testing sample set

with respect to the type I and type II faults can be obtained. A smaller value of IKL_i implies that the testing sample is similar to the corresponding training sample set in a statistic sense. Put another way, adding the testing sample into the training sample set causes a slight influence on the statistical distribution of the training sample set. Hence, the fault type of the testing sample can be discriminated. For instance, if $IKL_1 > IKL_2$, it can be concluded that the fault implied by the testing sample is more prone to the type II fault than type I fault. By this way, all the testing sample sets can be classified into one of the two fault types.

Following the same manner, the proposed method can be straightforwardly applied to a more general case in which the number of fault modes/damage levels to be classified is greater than two. In this case, the fault modes/damage level of the testing sample can be distinguished by looking for the smallest integrated KLID among all the known fault modes/damage levels.

3.2. Two case studies

The effectiveness of the proposed method in terms of diagnosing rotating machinery faults is validated in this section through two case studies of the bevel gears and the rolling element bearings.

3.2.1. Experimental rigs

Case 1: Experiments are performed on a machinery fault simulator produced by Spectra Quest, Inc. The experimental setup and the bevel gears to be tested are presented in **Figure 4**. The experimental setup composed of a motor, a coupling, bearings, two bevel gearboxes (one good right angle gearbox and one worn right angle gearbox), discs, belts, and a shaft. The bevel gearbox is driven by an AC motor and coupled with rub belts. The rotation speed was fixed to 1800 r/min. Three faulty gears, i.e., worn gear, gear with missing teeth, and gear with broken tooth, were simulated on the experimental setup. The raw vibration data were collected by an accelerometer that was mounted on the top of the gearbox. The data sampling rate was 20 kHz, and the data length was 4096 points [20].

Figure 4. The experiment rig and the four bevel gears with different damages. (a) Normal gear, (b) gear with broken tooth, (c) gear with missing teeth, and (d) gear with worn tooth.

Case 2: The experimental data are from the Case Western Reserve University [21]. The experimental rig is consisting of the Reliance Electric 2HP IQPreAlet, which is connected to a dynamometer. The bearings supporting the motor shaft were examined. Faults were artificially generated by creating crack size of 0.007, 0.014, 0.021, and 0.028 inches on the drive-end bearing through the electric discharge machining. These faults are separately distributed on the inner raceway, rolling element, and outer raceway. The raw vibration signals were collected by the two accelerometers mounted on the motor housing and the outer race of the drive-end bearing. The sampling frequency was set to be 12 kHz, whereas the sampling length was 12 k. The rotating speed was 1750 r/min. The detailed settings of this experiment can be found in Ref. [21].

3.2.2. Experimental testing and results

The raw data from the above two experimental setups are used to validate the proposed method. Without loss of generality, the data sets with the same type of defects or severity are randomly divided into training samples and testing samples. **Table 1** gives the training and testing sample sizes, the places of defects, and the defect sizes of the two case studies. The data set A in **Table 1** is from the Case 1, whereas the data sets B and C come from the Case 2. The capability of the proposed method in distinguishing the types of defects is examined through the data sets A and B. The capability of the method in identifying the severity of the same type of defect is validated through the data set C.

Data set		Number of training samples	Number of testing samples	Defect size (inch) (training/testing)*	Condition
A		35	35	–	Normal
		35	35	–	Broken tooth
		35	35	–	Missing teeth
		35	35	–	Worn tooth
B	B_1	35	35	0.007/0.021	Inner race
		35	35	0.007/0.021	Ball
	B_2	35	35	0.021/0.007	Inner race
		35	35	0.021/0.007	Ball
C		35	35	0.007	
		35	35	0.014	Inner race
		35	35	0.021	

*'-' for the data set A denotes the defect sizes of the training and testing samples are exactly the same but unmeasurable by a physical dimension.

Table 1. The data sets for defect and severity classification.

Two hundred and eighty data sets with four different operation conditions, i.e., normal condition, bevel gear with broken tooth, bevel gear with missing teeth, and bevel gear with worn tooth, are included in the data set A. The defect sizes of training and testing sample sets are exactly the same. It can be, thus, regarded as a four-class classification problem.

The data set B composed of 280 data sets of the faulty bearings has only two types of fault modes, i.e., inner race fault and ball fault. The data set B is divided into two subsets, i.e., subsets B1 and B2. Each of the subsets has 140 data samples. The study can be conducted to investigate the effectiveness of the proposed method if the fault mode of the training sample set is exactly the same as the testing sample set but the defect sizes are different. For the subset B1, 70 samples with the fault detect size of 0.007 inches are treated as the training sample set, whereas the remaining 70 samples with the fault detect size of 0.021 inches are the testing samples. To the contrary, for the subset B2, the training sample set of the subset B1 is treated as the testing set of the subset B2 and the testing set of the subset B1 is treated as the training sample set of the subset B2.

The data set C is consisting of 210 samples. The data set C is collected from the case where a defect is on the inner race. The defect sizes for the data set are 0.007, 0.021, and 0.028 inches. The aim of examining these three data sets is to validate the effectiveness of the proposed method in terms of identifying the defect severity (damage levels).

We exemplify the implementation of the proposed method to the data set A. Ninety-five time- and frequency-domain features are firstly extracted from the data set A. The effectiveness factors α_j of all the 95 features computed by the distance evaluation approach are shown in **Figure 5**, and the first 10 features with the greatest values are selected among all the 95 features.

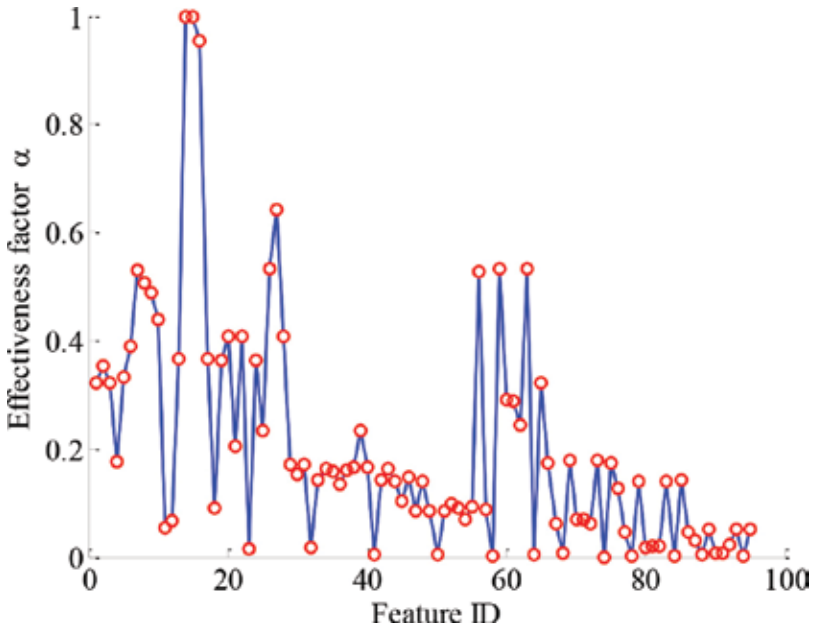


Figure 5. The effectiveness factor of all the extracted features.

Consequently, the probability density functions of the *j*th feature of the training sample sets for the four conditions, i.e., bevel gears with normal, broken tooth, missing teeth, and worn tooth

http://dx.doi.org/10.5772/67360

conditions respectively, can be obtained by the KDE, and they are denoted as $KD_i^j(i = 1,2,3,4)$. In the next step, a sample randomly picked up from the testing sample sets is included into the four training sets, and then, the corresponding probability density functions $TKD_i^j(i = 1,2,3,4)$ for the new sample set can be estimated. **Figures 6** and **7** give two examples of the probability density functions TKD_i^j of the first feature for the four training sample sets when a sample from the one of the four testing sampling sets is added.

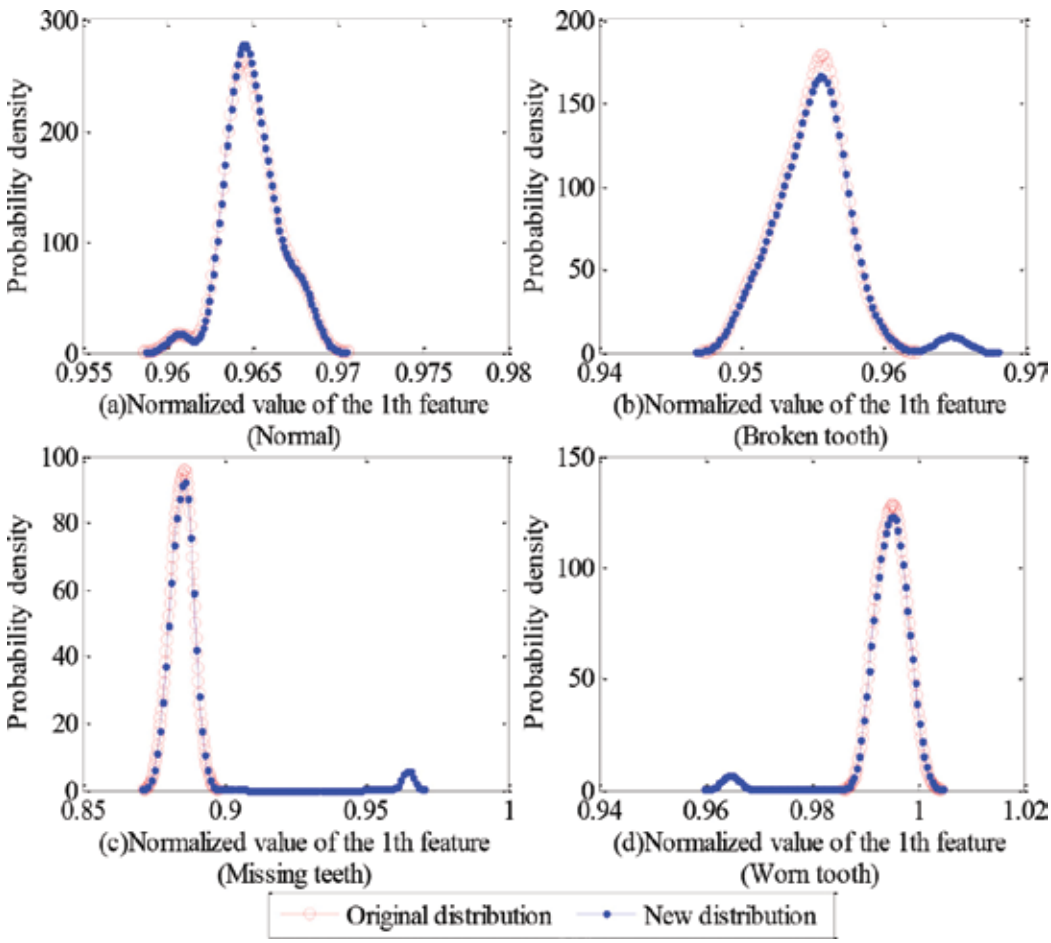


Figure 6. The original probability density functions of the four training sample sets and the corresponding new probability density functions after adding a normal testing sample.

In **Figures 6** and **7**, the red curves with circles are the original probability density functions of the first feature of the corresponding training set. The blue curves with dots represent the new probability density functions when a testing sample is included. For instance, as observed in **Figure 6(a)**, when a testing sample from the normal condition is added, the new probability density function of the first feature is almost the same as the original probability density function.

However, as seen from **Figure 6(b)–(d)**, the probability density functions exhibit a larger discrepancy when the testing sample from the normal condition is included to the other three training sample sets. Because the statistical characteristics of the first feature of the testing sample from the normal condition are distinct from these samples from the other three conditions, and the new probability density functions, therefore, generate a larger discrepancy from the original ones. Likewise, as observed in **Figure 7**, if the conditions of the new sample and the training sample sets (i.e., missing teeth) are the same, the new sample added to the training sample sets has minor impact on the probability density functions, otherwise a greater influence can be seen.

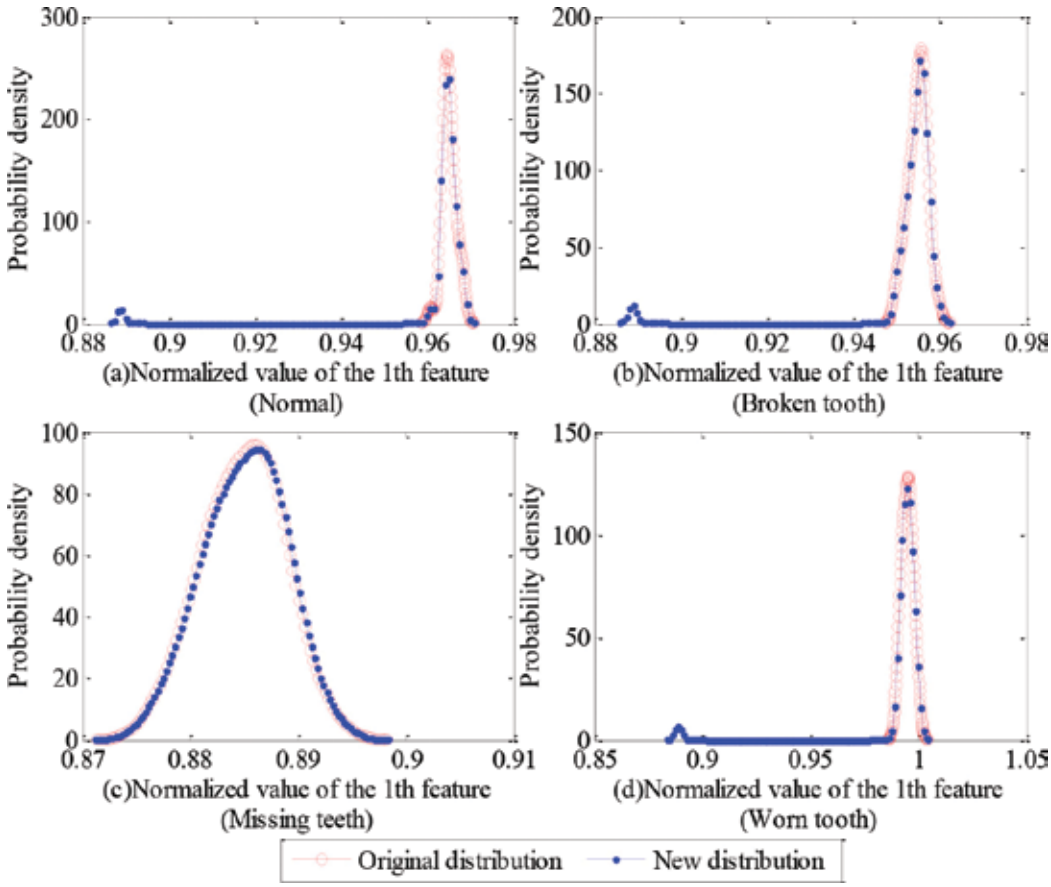


Figure 7. The original probability density functions of the four training sample sets and the corresponding probability density functions after adding a testing sample with missing teeth.

In the next step, the KLID is used to measure the difference between the original and the new distributions of the first feature in a quantitative way. The results are denoted as $KL_i^1 (i = 1,2,3,4)$ for the four different conditions. Following the same manner, the KLIDs can be evaluated for all the selected features. The integrated KLIDs, denoted as $IKL_i (i = 1,2,3,4)$, that aggregate the KLIDs of all the selected features are assessed based on the weights of the 10 selected features through Eq. (12).

The classification accuracy that is measured by evaluating the percentage of correctly distinguishing the fault modes or defect levels for the three data sets is presented in **Table 2**. A greater value of the percentage is favorable. The advantages of the proposed method are demonstrated through comparing the results from two conventional data-driven fault diagnosis methods, i.e., SVM-based fault diagnosis method and the back-propagation (BP) network-based fault diagnosis method. The parameter σ in SVM was optimized by the grid search method. The three-layer BP network was used, and it thresholds and weights were determined by the genetic algorithm to seek the global optimal solution. The results of the comparative study are presented in **Table 2**. For the data set A, the training and testing accuracy of the BP network-based fault diagnosis method are higher than those of the SVM-based fault diagnosis method. As opposed to the data set A, the SVM-based fault diagnosis method has a high training and testing accuracy for the data set B, whereas the BP network-based fault diagnosis method is inferior for the data set B and its accuracy is less than 90%. For the data set C, both the SVM-based fault diagnosis method and the BP network-based fault diagnosis method exhibit a relatively high accuracy. As seen from **Table 2**, the proposed method is superior to the two conventional methods on all the three data sets, and its accuracy reaches 100%.

Data set	SVM (%)		BP network (%)		The proposed method (%)	
	Training	Testing	Training	Testing	Training	Testing
A	95.25	92.14	**100**	99.62	**100**	**100**
B	**100**	98.10	89.29	87.14	**100**	**100**
C	**100**	97.86	**100**	96.19	**100**	**100**

Table 2. The classification accuracy of the three methods.

4. The proposed heath assessment approach

Following the similar framework as Section 3, the procedures of the proposed data-driven health assessment approach for rotating machinery are presented in **Figure 8**. Instead of using the distance-based feature selection method, which is a supervised feature selection approach and needs to set the number of the states to discriminate, the principle component analysis (PCA) as an unsupervised feature selection tool is used here. The KDE and KLID are used together to construct a new health indicator, reflecting the health condition of the monitored rotating machinery.

Figure 8. The procedures of the proposed health assessment approach for rotating machinery.

4.1. The principle component analysis (PCA)

The PCA proposed by Pearson [22] is a statistical procedure aiming to extract the directions with strong variability in a data set, and it can convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. As its good capability in terms of reducing the dimensionality of data set, the PCA has been extensively used to deal with multivariate data in the field of pattern recognition, image processing, etc.

Mathematically, given a set of p dimensional feature vectors $\mathbf{x}_i (i = 1, 2, \ldots, n)$, the corresponding covariance matrix of the feature vectors can be computed by:

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}}, \tag{13}$$

where

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_i \tag{14}$$

The principal components (PCs) can be computed by solving the eigenvalue and eigenvector of the covariance matrix $\mathbf{C}$ as follows:

$$\mathbf{C}\mathbf{v} = \boldsymbol{\lambda}\mathbf{v} \tag{15}$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_p]$ are the eigenvalues of the covariance matrix $\mathbf{C}$ in a descending order and $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p]$ are the associated eigenvectors.

To represent the original feature vector through a lower dimensional feature vector, the first m ($m \le p$) eigenvectors that correspond to the first k largest eigenvalues will be selected. Oftentimes, a pre-specified threshold $\theta (\theta \in [0,1])$ needs to be given for a particular problem by user to satisfy:

$$\sum_{i=1}^{m}\lambda_i \Big/ \sum_{i=1}^{p}\lambda_i \ge \theta \tag{16}$$

A greater value of θ indicates to maintain a higher accuracy of the original feature vectors, and thus, more eigenvectors will be included. By this way, the number of eigenvectors for a particular problem can be determined so as to maintain the desired accuracy. The m dimensional feature vectors can be formulated as:

$$\boldsymbol{\varepsilon}_j = \mathbf{V}^{\mathrm{T}}\mathbf{x}_j, j = 1, 2, \ldots, n \tag{17}$$

By using the PCA, the dimensionality of the original feature vectors can be significantly reduced. The importance of the new feature vectors is denoted as $\mathbf{F} = (F_1, F_2, \ldots, F_m)$, and the importance of each new feature i can be evaluated by:

http://dx.doi.org/10.5772/67360

$$F_i = \lambda_i / \sum_{j=1}^{m} \lambda_j, \quad (i = 1, 2, \cdots, m) \tag{18}$$

4.2. The procedure of the proposed health assessment approach

The key idea behind the proposed health assessment approach is that the statistical characteristics of the samples at the good condition would exhibit an apparent discrepancy with that of the samples at the abnormal condition. In our work, the statistical characteristics of the samples are characterized by the KDE, whereas the KLID provides a quantitative way to measure the statistical discrepancy of the online monitoring samples with respect to the reference samples that are collected when the monitored device is at its good condition.

The overall flowchart of the proposed health assessment approach is shown in **Figure 9**. As shown in **Figure 9**, the features sensitive to the health status of rotating machinery are chosen. By conducting the PCA, the dimensionality of the selected features can be further reduced so as to reduce the computational burden in the ensuing steps. A moving window with width k will be used to dynamically construct a set of samples to evaluate the health condition of the monitored rotating machinery. An illustration of constructing sample sets over time through the moving window is delineated in **Figure 10**. With the assumption that the rotating machinery is at its good condition at the early stage of use, the samples collected by the moving window at the beginning of use will serve as the reference samples, whereas the samples collected by the moving window at the later stage will be statistically compared with the reference samples.

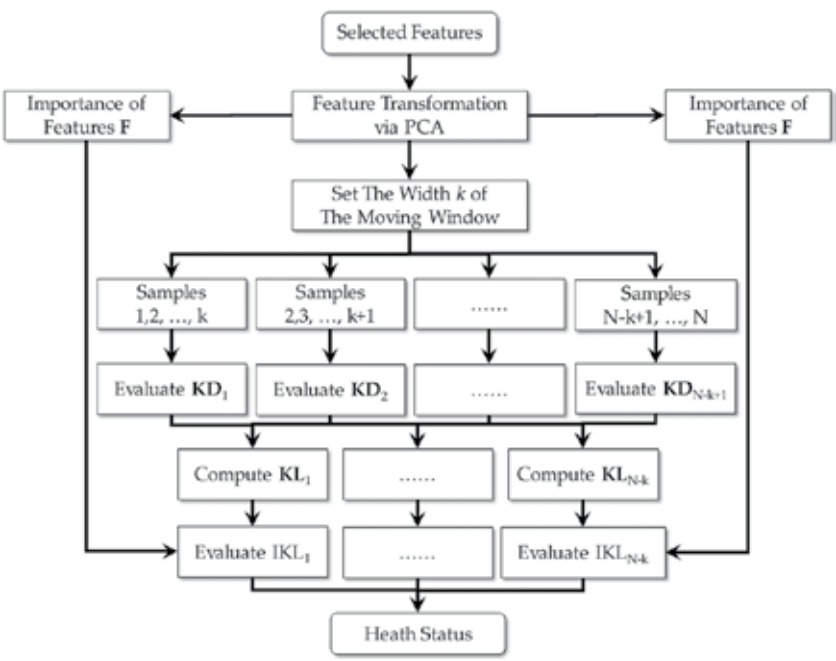


Figure 9. The flowchart of the proposed health assessment approach.

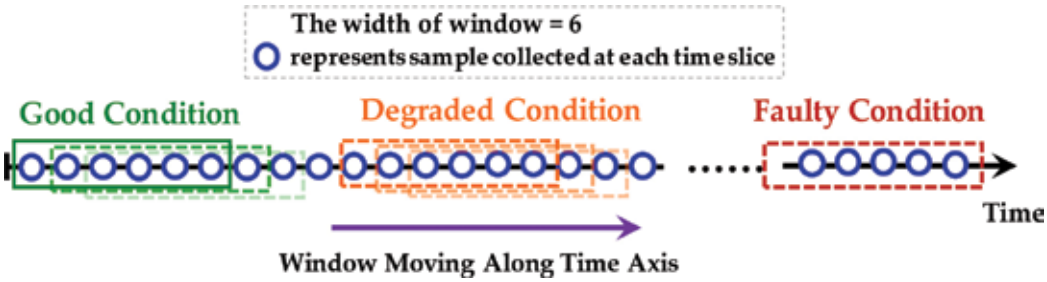


Figure 10. An illustration of the *k*-width moving window.

The statistical characteristic of each sample set in the moving window is characterized by the kernel density function, and it is denoted as KD_i^j for the *j*th ($j = 1, 2, \ldots, m$) PCA feature at the *i*th ($i = 1, 2, \ldots, N - k + 1$) window. $\mathbf{KD}_i = \left(KD_i^1, KD_i^2, \cdots, KD_i^m\right)$ is a collection of all the PCA features at the *i*th ($i = 1, 2, \ldots, N - k + 1$) window. The statistical discrepancy between the samples at the *i*th ($i = 1, 2, \ldots, N - k + 1$) window and $\mathbf{KD}_1 = \left(KD_1^1, KD_1^2, \cdots, KD_1^m\right)$ is quantified by the KLID and denoted as $\mathbf{KL}_i = \left(KL_i^1, KL_i^2, \cdots, KL_i^m\right)$ where KL_i^j computed by Eq. (10) is the KLID of the *j*th ($j = 1, 2, \ldots, m$) PCA feature at the *i*th ($i = 1, 2, \ldots, N - k + 1$) window. By taking into account the importance of the PCA features, the integrated KLID, denoted as IKL_i ($i = 1, 2, \ldots, N - k$), can be evaluated by Eq. (12), where $F_i (i = 1, 2, \ldots, m)$ in Eq. (12) takes values from Eq. (18). It should be noted that in the proposed approach IKL_i act as the health indicator for rotating machinery. A smaller value of IKL_i represents that the condition of the monitored device is close to the normal condition. On the contrary, if the condition of the monitored device gradually deviates the normal condition due to defects or faults, IKL_i will become a greater value.

4.3. A case study

To validate the effectiveness of the proposed method in terms of assessing the health status of rotating machinery, a case study for rolling element bearing is presented in this section.

4.3.1. Experimental setup

The experimental data are from the intelligent maintenance system (IMS) at the University of Cincinnati [23]. The run-to-failure data were collected from the experimental rig as shown in **Figure 11**, where the rolling bearings were working under a constant load condition. The rolling bearing test rig hosts four test Rexnord ZA-2115 double row rolling bearings on one shaft. Each row of the rolling bearings has 16 rollers, the section diameter is 71.5 mm, the rolling diameter is 8.4 mm, and the contacting angle of the roller is 15.17°. The rotation speed was set to be 2000 rpm. The sampling rate was 20 kHz, whereas the data length was 20,480 points. Three testing (i.e., Testings 1, 2, and 3) with identical rolling bearings were executed on this experimental rig.

In the reported experiment, three run-to-failure data sets were collected. At the end of Testing 1, an inner race defect occurred on Bearing 3. Bearing 4 developed a roller defect. At the end of

http://dx.doi.org/10.5772/67360

Testing 2, an outer race defect was found on Bearing 1. At the end of Testing 3, an outer race defect happened on Bearing 3. In our study, the run-to-failure data sets from Testings 1 and 2 are used to validate our proposed approach.

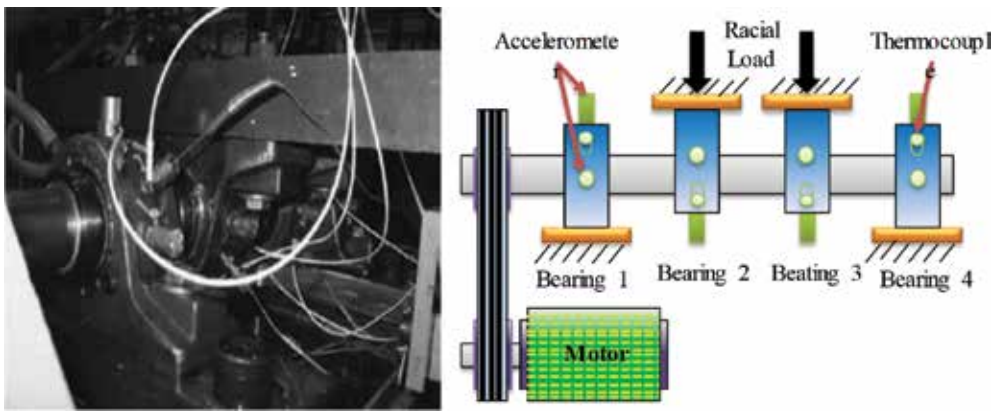


Figure 11. The rolling bearing experimental rig [23].

4.3.2. Results and analysis

In our study, the tools, like the ensemble empirical mode decomposition (EEMD), were first used to extract 95 representative features from the raw data sets. These features were transformed by the PCA to reduce the dimensionality. The importance of the principle components for Bearing 3 in Testing 1 and Bearing 1 in Testing 2 is shown in **Figure 12(a)** and **(b)**.

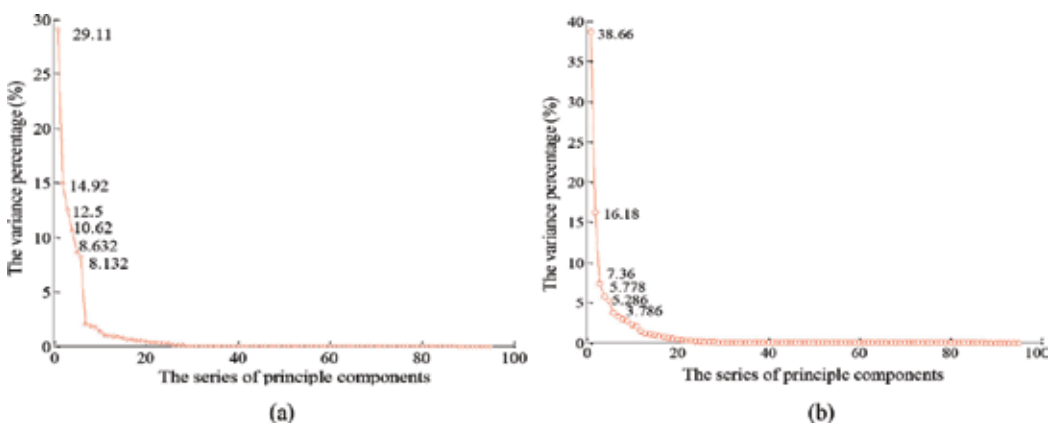


Figure 12. The importance of the principle components. (a) Bearing 3 in Testing 1 and (b) Bearing 1 in Testing 2.

From **Figure 12**, one can see that by using the first 10 principal components only, 90% accuracy can be maintained. On the other hand, with such a small sacrifice of accuracy, the dimensionality of features can be dramatically reduced from 95 to 10. Therefore, the first 10 principal

components were used as selected features and put into the proposed health assessment approach. The integrated KLID values from the proposed approach are plotted in **Figure 12 (a)** and **(b)** for Bearing 3 in Testing 1 and Bearing 1 in Testing 2, respectively, and these curves acting as the health indicator reflect the health status of the monitored bearings.

As shown in **Figure 13(a)**, the health indicator of Bearing 3 in Testing 1 has slight fluctuations at the early stage of the experiment. At the 1750th point, the health indicator rose up steeply and reached a great value in a short period of time (at the 1800th point). Such observation could indicate that the bearing put into use had a small manufacturing defects or a slight damage. It, therefore, leaded to the slight fluctuations of the health status at the beginning of the experiment. The tiny defect became serious suddenly at the 1750th points.

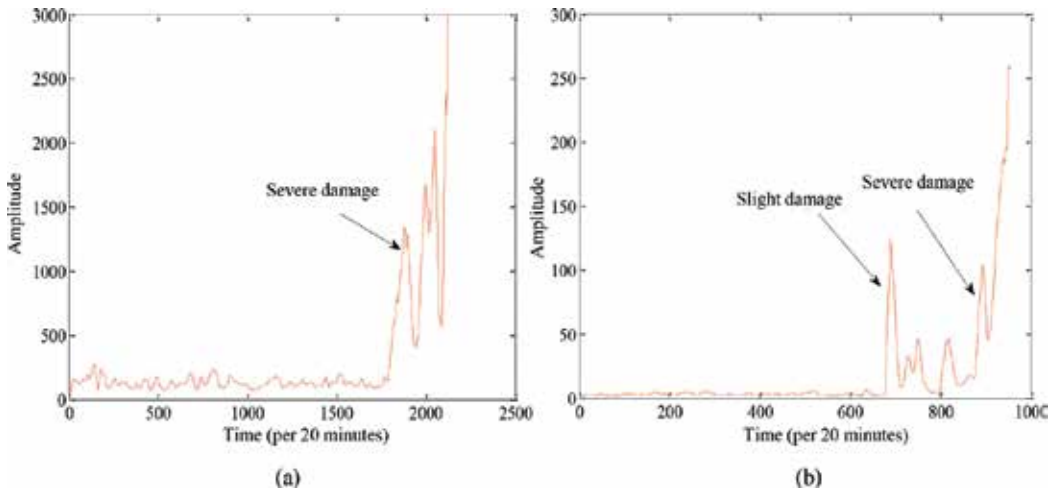


Figure 13. The results of health assessment. (a) Bearing 3 in Testing 1 and (b) Bearing 1 in Testing 2.

In **Figure 13(b)**, the health indicator of Bearing 1 in Testing 2 experienced two stages, namely the slight damage stage and the severe damage stage. Due to the slight damage, the health curve went up at the 680th point and reached the first peak value around the 700th point. However, as the bearing entered a new unhealthy but stable state, the health status of the bearing has a slight improvement as one can observed that the health curve dropped down for a while. Two hundred points (around 70 hours) later, the bearing jumped into the severe damage stage as shown in **Figure 13(b)**.

To illustrate the effectiveness of the proposed approach, the results from a recent literature are compared. The Locality Preserving Projection and Gaussian mixture model (GMM) was developed in Ref. [24] to construct a health assessment model for Bearing 1 in Testing 2. As found in Ref. [24], the health curve changed after the 700th point, indicating the occurrence of a slight damage. However, our proposed health indicator which rose up at the 680th point and reached the first maximum value at the 700th point. To examine the status of Bearing 1 in Testing 2 at the 680th point, the empirical mode decomposition (EMD) [25] was used to decompose the collected vibration data to four levels. The Hilbert transformation (HT) was

http://dx.doi.org/10.5772/67360

performed on the four intrinsic mode functions (IMFs). For more details about HT, please refer to Ref. [20]. The corresponding results are shown in **Figure 14**. From **Figure 14(b)**, one can easily find the frequency component of 236.3 Hz and its 2–4 times frequency components. On the other hand, the ball bass frequency at outer race (BPFO) can be theoretically calculated as follows:

$$f_{BPFO} = \frac{1}{2} \times n \times [1 - (d/D) \times \cos\varphi] \times N \quad = 236.4\text{Hz} \tag{19}$$

Thereby, one can conclude that the outer race defect occurred at the 680th point. Furthermore, this result illustrates that the proposed health assessment approach has a better capability to detect the incipient defect than the method proposed in Ref. [24].

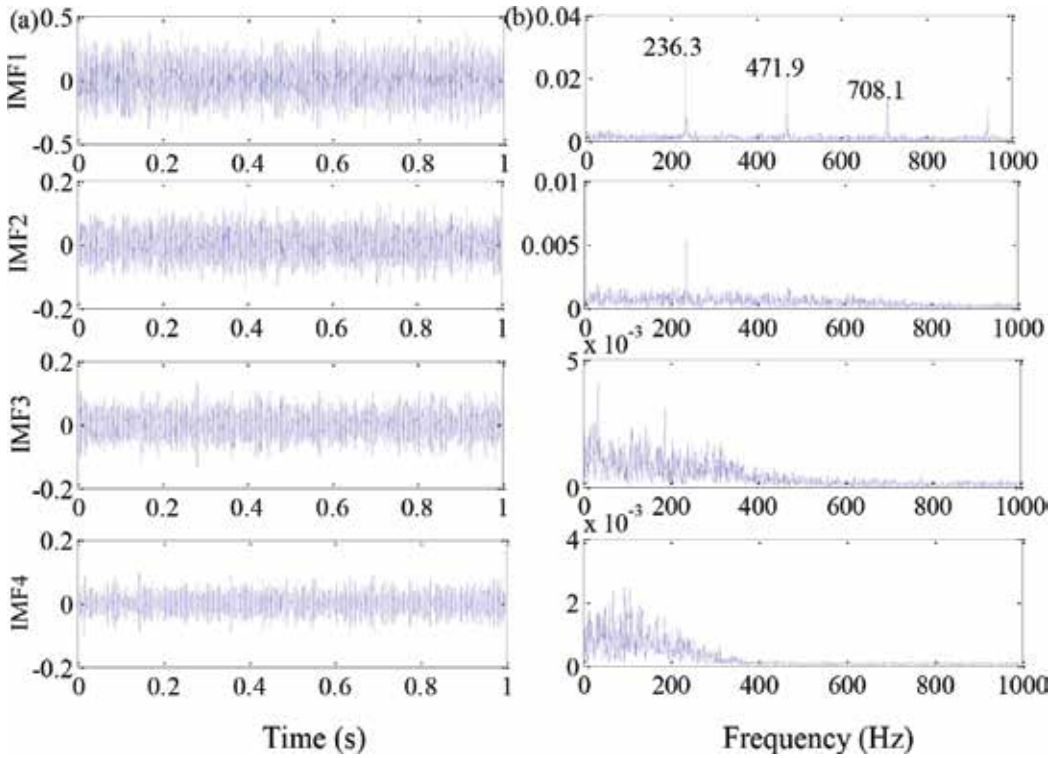


Figure 14. The results of the EMD and the corresponding Hilbert spectrum of Bearing 1 in Testing 2 at the 680th point. (a) The first four IMFs and (b) the Hilbert spectrum.

5. Conclusion

In this chapter, based on the kernel density estimation and the Kullback-Leibler divergence, a new data-driven fault diagnosis approach and a new health assessment approach are developed

by examining the statistical characteristics of the collected sample sets. By using the KDE and the KLID, the fault types or health status can be identified by comparing the integrated KLID of selected features. As demonstrated in the fault diagnosis examples, the proposed fault diagnosis approach has an exceptional performance on faulty pattern recognition, and it outperforms the conventional SVM-based and BP network-based methods. Meanwhile, in the example of health assessment, the proposed health assessment approach, which takes account of the statistical characteristics of sample sets, is capable of quantitatively tracking the health condition of the monitored rotating machinery.

Acknowledgement

The authors greatly acknowledge the grant support from the National Natural Science Foundation of China under contract number 71371042 and the Fundamental Research Funds for the Central Universities under contract number ZYGX2015J082.

Author details

Yu Liu*, Chen-Yao Yan and Fan Zhang

*Address all correspondence to: yuliu@uestc.edu.cn

School of Mechatronics Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, P.R. China

References

[1] Jardine A.K.S., Lin D., Banjevic D.. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing. 2006;**20**(7):1483 –1510.

[2] Heng A., Zhang S., Tan A.C.C., Mathew J.. Rotating machinery prognostics: state of the art, challenges and opportunities. Mechanical Systems and Signal Processing. 2009;**23**(3):724–739.

[3] Chen J., Patton R.J.. Robust Model-Based Fault Diagnosis for Dynamic Systems. New York: Springer; 1999.

[4] Sekhar A.S.. Model-based identification of two cracks in a rotor system. Mechanical Systems and Signal Processing. 2004;**18**(4):977–983.

[5] Li Y., Kurfess T.R., Liang S.Y.. Stochastic prognostics for rolling element bearings. Mechanical Systems and Signal Processing. 2000;**14**(5):747–762.

http://dx.doi.org/10.5772/67360

[6] Tibaduiza D.A., Torres-Arredondo M.A., Mujica L.E., et al. A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring. Mechanical Systems and Signal Processing. 2013;**41**(1):467–484.

[7] Lei Y.G., He Z.J., Zi Y.Y.. A new approach to intelligent fault diagnosis of rotating machinery. Expert Systems with Applications. 2008;**35**(4):1593–1600.

[8] Cui H., Zhang L., Kang R., Lan X.. Research on fault diagnosis for reciprocating compressor valve using information entropy and SVM method. Journal of Loss Prevention in the Process Industries. 2009;**22**(6):864–867.

[9] Saravanan N., Ramachandran K.I.. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). Expert Systems with Applications. 2010;**37**(6):4168–4181.

[10] Widodo A., Yang B.S.. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing. 2007;**21**(6):2560–2574.

[11] Rafiee J., Arvani F., Harifi A., Sadeghi M.H.. Intelligent condition monitoring of a gearbox using artificial neural network. Mechanical Systems and Signal Processing. 2007;**21**(4):1746–1754.

[12] Zhao F., Tian Z., Zeng Y.. Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method. IEEE Transactions on Reliability. 2013;**62**(1):146–159.

[13] Yang B.S., Kim K.J.. Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals. Mechanical Systems and Signal Processing. 2006;**20**(2):403–420.

[14] Rosenblatt M.. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics. 1956;**27**(3):832–837.

[15] Parzen E.. On estimation of a probability density function and mode. The Annals of Mathematical Statistics. 1962;**33**(3):1065–1076.

[16] Kullback S., Leibler R.A.. On information and sufficiency. The Annals of Mathematical Statistics. 1951;**22**(1):79–86.

[17] Wu Z.H., Huang N.E.. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis. 2009;**1**(1):1–41.

[18] Lei Y.G., He Z.J., Zi Y.Y.. Application of the EEMD method to rotor fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing. 2009;**23**(4):1327–1338.

[19] Yang B.S., Han T., An J.L.. ART–KOHONEN neural network for fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing. 2004;**18**(3):645-657. Epanechnikov V.A. Non-parametric estimation of a multivariate probability density. Theory of Probability & Its Applications. 1969;**14**(1):153–158.

[20] Zhang F., Liu Y., Chen C., Li Y., Huang H.Z.. Fault diagnosis of rotating machinery based on kernel density estimation and Kullback-Leibler divergence. Journal of Mechanical Science and Technology. 2014;**28**(11):4441–4454.

[21] Bearing Data Center, Case Western Reserve University. Available from: http://www.eecs.cwru.edu/laboratory/Bearing [Accessed: June, 2009].

[22] Pearson K.. On lines and planes of closest fit to systems of points in space. Philosophical Magazine. 1901;**2**(11):559–572.

[23] Lee J., Qiu H., Yu J., Lin J.. Rexnord Technical Services. Bearing Data Set [Internet]. Available from: http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/ [Accessed: March 2014]

[24] Yu J.. Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models. Mechanical Systems and Signal Processing. 2011;**25**(7):2573–2588.

[25] Huang N.E., Shen Z., Long S.R., et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London 1971, **454**:903–995.

Chapter 7

Dynamics-Based Vibration Signal Modeling for Tooth Fault Diagnosis of Planetary Gearboxes

Xihui Liang, Ming J. Zuo and Wenhua Chen

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67529

Abstract

Vibration analysis has been widely used to diagnose gear tooth fault inside a planetary gearbox. However, the vibration characteristics of a planetary gearbox are very complicated. Inside a planetary gearbox, there are multiple vibration sources as several sun-planet gear pairs, and several ring-planet gear pairs are meshing simultaneously. In addition, due to the rotation of the carrier, distance varies between vibration sources and a transducer installed on the planetary gearbox housing. Dynamics-based vibration signal modeling techniques can simulate the vibration signals of a planetary gearbox and reveal the signal generation mechanism and fault features effectively. However, these techniques are basically in the theoretical development stage. Comprehensive experimental validations are required for their future applications in real systems. This chapter describes the methodologies related to vibration signal modeling of a planetary gear set for gear tooth damage diagnosis. The main contents include gear mesh stiffness evaluation, gear tooth crack modeling, dynamic modeling of a planetary gear set, vibration source modeling, modeling of transmission path effect due to the rotation of the carrier, sensor perceived vibration signal modeling, and vibration signal decomposition techniques. The methods presented in this chapter can help understand the vibration properties of planetary gearboxes and give insights into developing new signal processing methods for gear tooth damage diagnosis.

Keywords: dynamic simulation, effect of transmission path, gear mesh stiffness, signal decomposition, fault diagnosis

1. Introduction

Planetary gearboxes are widely used in military and industrial applications. For example, they are main transmission components in military helicopters, wind turbines, and mining

trucks as shown in **Figure 1**. Comparing with fixed-shaft gearboxes, planetary gearboxes can afford higher torque load due to the load sharing among multiple gear pairs and generate larger transmission ratio with equal or smaller volume.

Figure 2 illustrates the structure of a one-stage planetary gear set that is composed of a sun gear, a ring gear, a carrier, and several planet gears. This planetary gear set can achieve multiple transmission scenarios as illustrated in **Table 1**. One transmission scenario can be selected in real applications based on individual application requirements. A planetary gear set is much more versatile in transmission scenarios comparing with a fixed-shaft gear set.

However, the versatility of planetary gearbox transmissions comes at a price: planetary gear sets are much more complicated to design and analyze. In the afternoon of April 1, 2009, a helicopter crashed into the North Sea and all 16 crew members died [1]. Later analysis concluded that this accident was caused by gear fatigue crack. Therefore, it is crucial to be able to detect early fault of planetary gearboxes; otherwise, large economic losses or catastrophes may occur. Vibration analysis, acoustic analysis, oil debris analysis, temperature analysis, and strain analysis are common techniques in the condition monitoring of gearbox systems. In this chapter, we only focus on the study of vibration analysis. Vibration analysis relies on the analysis of vibration signals to detect faults in a planetary gearbox.

There are mainly two ways to detect planetary gearbox damages based on vibration analysis. One way is to physically measure vibration responses of planetary gearboxes using vibration sensors and then analyze these vibration signals using advanced signal processing techniques to determine the health condition of planetary gearboxes [2–6]. **Figure 3** gives an example which is a planetary gearbox test rig located in the Department of Mechanical Engineering of the University of Alberta. Four types of experiments were performed in this test rig: gear tooth crack experiments, tooth pitting experiments, run-to-failure experiments, and various load and speed experiments [2]. Planetary gear fault detection/classification techniques were developed [3–6] by analyzing vibration signals collected from this experimental test rig. The second way is to model system responses (vibration signals) of planetary gearboxes using physical laws, like the Newton's laws of motion and then analyze simulated vibration signals aiming to reveal the nature of fault symptoms. The simulated vibration signals do not have the environmental noise interference. They can reveal fault features more easily than the physically measured signals. However, environmental noise did exist in real applications. The fault features revealed by the simulated signals may be submerged by the noise and

Figure 1. Applications of planetary gearboxes: (a) helicopter, (b) wind turbine, and (c) mining truck.

http://dx.doi.org/10.5772/67529

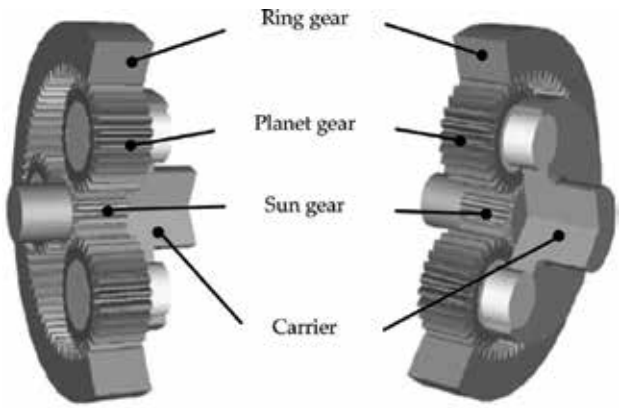


Figure 2. Structure of a planetary gear set.

Option	Inputs	Output
1	Sun (carrier fixed)	Ring
2	Ring (carrier fixed)	Sun
3	Carrier (sun fixed)	Ring
4	Ring (sun fixed)	Carrier
5	Sun (ring fixed)	Carrier
6	Carrier (ring fixed)	Sun
7	Sun and carrier	Ring
8	Ring and sun	Carrier
9	Ring and carrier	Sun

Table 1 Transmission scenarios of a planetary gear set as given in **Figure 2**.

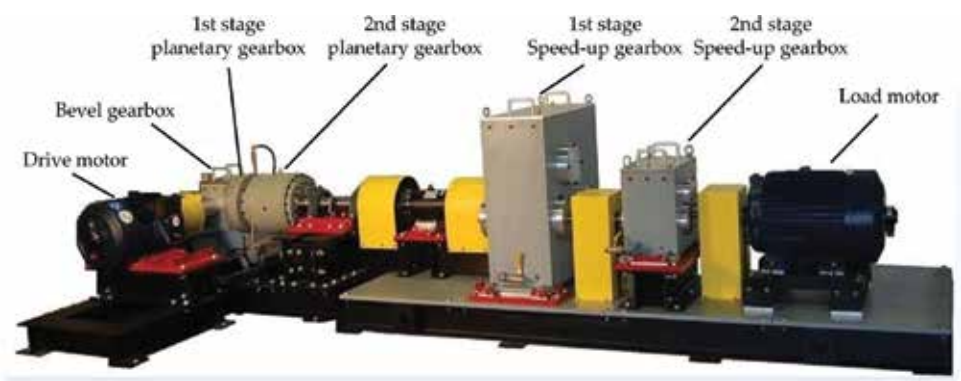


Figure 3. A planetary gearbox test rig located in the University of Alberta.

become ineffective. Therefore, environmental noise effect should be considered before applying the fault detection techniques developed using the simulated signals. In this chapter, we limit our focus on the second way of planetary gearbox modeling and fault diagnosis.

Three types of system responses can be obtained from analytical or numerical modeling of planetary gearboxes. The first type analysis mainly focuses on gear nature frequency and mode analysis [7–12]. The second type analysis is to analyze vibration properties of individual gears or dynamic forces of a gear pair among a planetary gear set [13–27]. In the first two types of analysis, the effect of gear transmission path is not considered. **Figure 4** illustrates three possible transmission paths for a vibration induced by a sun-planet meshing. Due to the effect of transmission path, some vibrations may be attenuated or submerged in the process of reaching a transducer located on the gearbox housing. Some researchers modeled the resultant vibration signals sensed by a transducer by considering multiple vibration sources inside a planetary gearbox and the effect of transmission path [5, 28–33]. A planetary gearbox has multiple vibration sources due to several sun-planet gear pairs and ring-planet gear pairs are meshing simultaneously. The effect of transmission path is mainly induced by the rotation of the carrier that causes the varying distance between a planet gear and a transducer mounted on the gearbox housing. The resultant vibration signals are the third type of system response. In the real applications, we generally install transducers on gearbox housing or bearing housing to collect vibration signals. Multiple vibration sources go through different transmission paths and reach the transducer. Therefore, it is important to consider both multiple vibration sources and the transmission path effect in vibration signal modeling. In this chapter, we focus on the modeling of resultant vibration signals of a planetary gearbox. An improved lumped parameter model [31] is used to simulate gearbox vibration source signals. This model is similar to the one reported in Ref. [7] with three distinctions: (1) the planet deflections are described in the horizontal and vertical coordinates, (2) both the gyroscopic force and the centrifugal force are incorporated, and (3) more accurate physical parameters are adopted. In addition, a modified Hamming function is used to model the effect of transmission path. This model is an improvement of widely used Hamming model [5, 28, 29]. The procedures to obtain resultant vibration signals are summarized in **Figure 5**. In the end, a new signal decomposition technique is used to enlarge gearbox fault signatures based on analysis of simulated vibration signals.

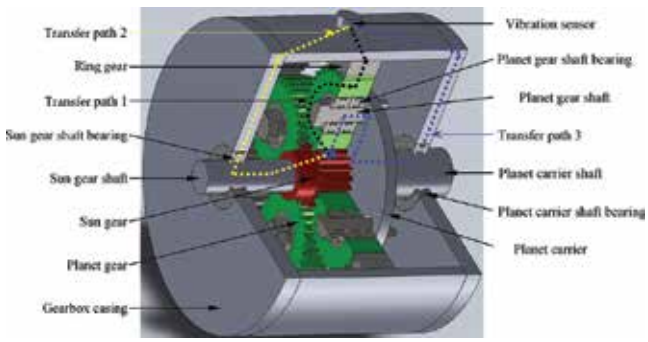


Figure 4. Transmission paths [5].

http://dx.doi.org/10.5772/67529

Figure 5. Procedures to obtain resultant vibration signals.

In this chapter, we present detailed procedures of modeling resultant vibration signals of a planetary gearbox with tooth damages and vibration signal decomposition techniques for fault diagnosis. Some challenges are described and analyzed. The remaining part of this chapter is organized as follows. In Section 2, a dynamic model is illustrated to simulate vibration source signals of a planetary gearbox. In Section 3, the methods to evaluate model parameters are described. In Section 4, the modeling of transmission path effect is given. Section 5 presents simulated vibration signals and fault symptoms. Section 6 describes vibration signal decomposition techniques for gear tooth damage diagnosis. Section 7 draws a summary and points out future research topics in dynamics-based vibration signal modeling and fault diagnosis of planetary gearboxes.

2. Dynamic modeling for simulation of vibration source signals

Planetary gearbox transmission systems are complex, and it is hard to consider all details of the transmission. As a result, people simplify the problem as much as possible while retaining all of the important and relevant features. A gearbox transmission can be modeled as a lumped system or a distributed system. A lumped system is simpler than a distributed system. A distributed system may be able to cover more details of a gearbox transmission, but its governing equations are hard to solve. The dependent variables of a lumped system are functions of time alone, and in general, the equation of motions is represented by ordinary differential equations. By contrast, the dependent variables of a distributed system are functions of time and one or more spatial variables, and the equation of motions can only be expressed by partial differential equations. Generally, the more details covered for a gearbox transmission system, the more complicated the equation of motions. Analytical solutions to complicated differential equations are hard to obtain. Numerical methods are mostly applied to solve them. However, four types of errors may be induced using numerical methods: round-off error, truncation error, accumulated error, and relative error [34]. It is time-consuming to solve complicated differential equations and sometimes it is even impossible to get a proper solution. Therefore, it is a trade-off between details to be covered and computation difficulties.

Figure 6 gives a typical modeling of a planetary gear set [30]. It is a 2D lumped parameter model. Each gear has three degrees of freedoms: angular rotation and transverse motions in the x- and y-directions. The gear mesh interface is modeled as a spring-damper system. Each bearing is also modeled as a spring-damper system. Other practical phenomena such as gear transmission error, backlash, tooth friction, gear shaft effect, gearbox housing effect, and gear misalignments are not considered in that model, but these can be selectively supplemented

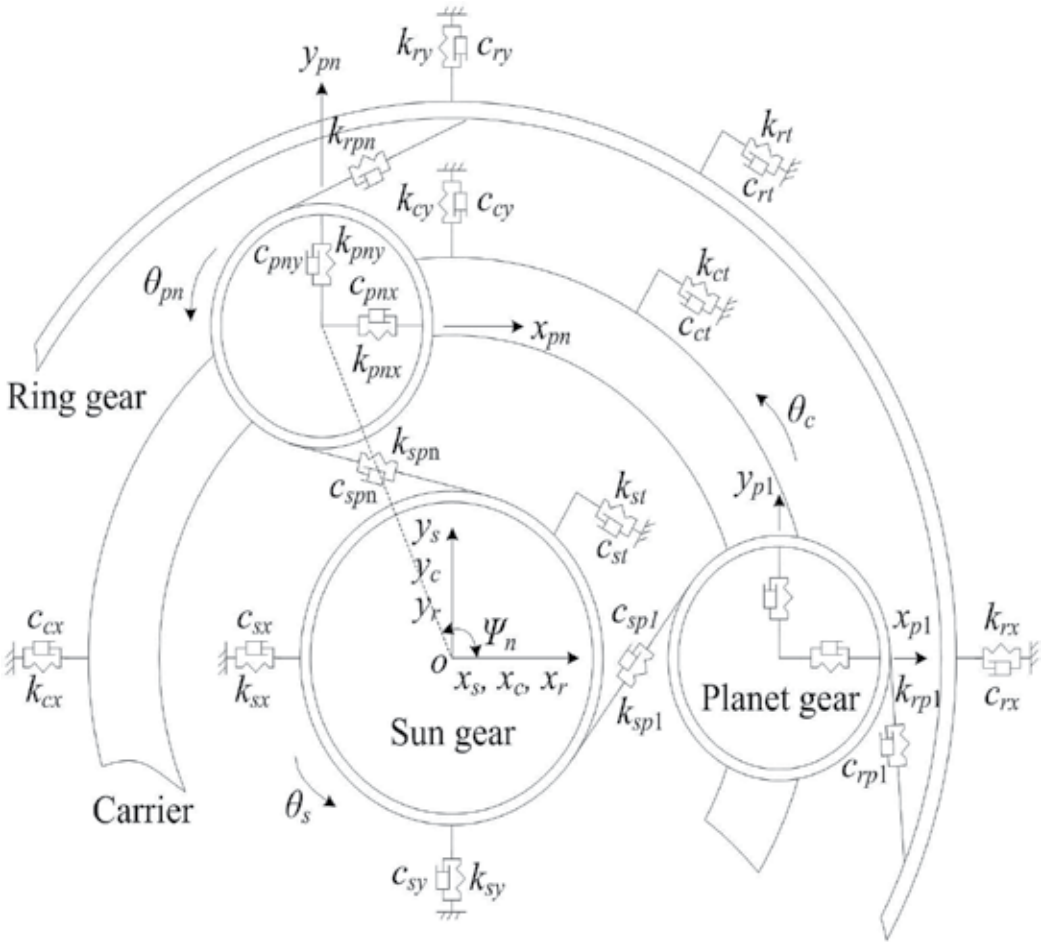


Figure 6. Modeling of a planetary gear set [30].

according to individual future research of interest. The model reported in Ref. [30] is used directly in this chapter to simulate vibration source signals of a planetary gearbox. The governing equations for this model are available in Ref. [30]. The vibration source signals will be illustrated later in Section 5.

3. Evaluation of physical parameters

For the dynamic model given in **Figure 6**, governing equations are available in Ref. [30]. But, several physical parameters in this model need to be evaluated before solving these equations, for example, gear mesh stiffness and damping, bearing stiffness and damping, and gear moment of inertia.

http://dx.doi.org/10.5772/67529

The most important one physical parameter is gear mesh stiffness. There are basically two ways to evaluate gear mesh stiffness: finite element method (FEM) and analytical method (AM). FEM is flexible to model any shaped gear and gear fault, but it is sensitive to contact tolerances, mesh density, and the type of finite elements selected. As the increase of mesh density, the numerical accuracy is improved, but the computation cost goes up. To reduce computation cost of FEM, Parker et al. [35] and Ambarisha and Parker [8] developed a combined element/contact mechanics model in gear mesh stiffness evaluation. Liang et al. [36] used linear finite element analysis to save computation cost in gear mesh stiffness evaluation. AM assumes a gear tooth as a nonuniform cantilever beam and beam theories are applied to evaluate gear mesh stiffness. AM has higher computational efficiency than FEM. But, AM is hard to model shape-complicated gear teeth and some gear faults. Potential energy method is one popular AM. This method has been used to evaluate mesh stiffness of perfect gears [37, 38], gears with crack [39–45], gears with a single tooth pit/spalling [46, 47], gears with multiple tooth pits [36], gears with a chipped tooth [39], gears with tooth plastic inclination deformation [19], gears with tooth profile modification [48], and gears with carrier misalignment errors [22].

In this section, the potential energy method used in Ref. [43] is illustrated to evaluate the mesh stiffness of a planetary gear set in healthy and cracked tooth conditions. The gear tooth is modeled as a nonuniform cantilever beam starting from gear base circle. The total energy stored in a pair of meshing gears is considered to be the summation of Hertzian contact energy, bending energy, shear energy, and axial compressive energy that corresponds to Hertzian contact stiffness, bending stiffness, shear stiffness, and axial compressive stiffness, respectively. When a gear tooth crack occurs, the effective tooth length, the area, and area moment of tooth sections of a cracked tooth are different from that of a perfect tooth, which leads to the gear mesh stiffness reduction. In Ref. [43], the gear tooth crack path is modeled in a straight line shape starting from the critical area of tooth root as shown in **Figure 7**. The critical area is around the maximum principle stress point at the tooth root. The equations for Hertzian contact stiffness (k_h), bending stiffness (k_b), shear stiffness (k_s), and axial compressive stiffness (k_a) are derived and available in Ref. [43]. These equations are expressed as a function of gear rotation angle (given gear geometry, material information, and crack information). Users can use these equations directly to evaluate gear mesh stiffness even though they are not familiar with beam and/or gear meshing theories. The total effective mesh stiffness k_t can be obtained as follows:

$$k_t = \sum_{i=1}^{m} \frac{1}{\frac{1}{k_{h,i}} + \frac{1}{k_{b1,i}} + \frac{1}{k_{s1,i}} + \frac{1}{k_{a1,i}} + \frac{1}{k_{b2,i}} + \frac{1}{k_{s2,i}} + \frac{1}{k_{a2,i}}} \tag{1}$$

where i represents the ith pair of meshing teeth and the subscripts 1 and 2 denote the driving gear and the driven gear, respectively.

Figure 8 summarizes the steps to obtain mesh stiffness of a planetary gear set. First, the mesh stiffness of a pair of sun-planet gears (a pair of external gears) and a pair of ring-planet gears (a pair of internal gears) should be evaluated using the potential energy method, respectively. Then, by incorporating gear mesh phase relationships [49], the mesh stiffness of other sun-planet gear pairs and ring-planet gear pairs can be obtained.

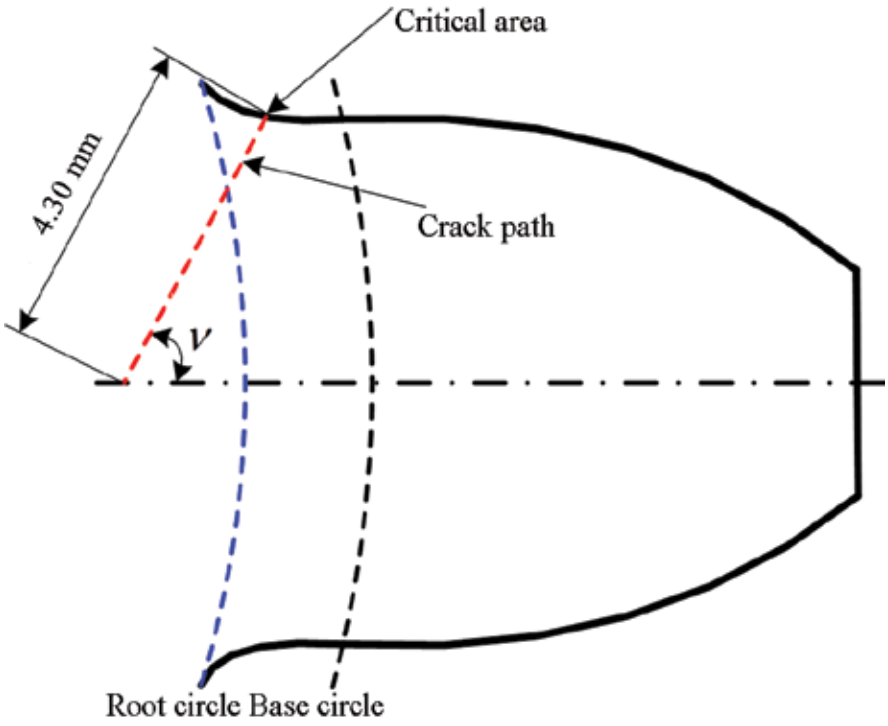


Figure 7. Tooth crack modeling.

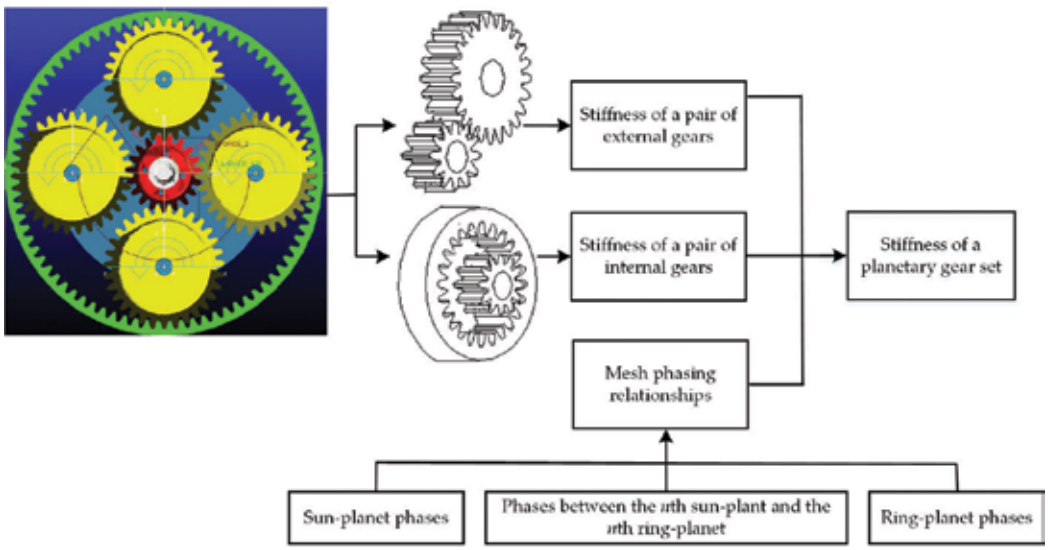


Figure 8. Steps to obtain mesh stiffness of a planetary gear set [31].

http://dx.doi.org/10.5772/67529

Table 2 gives the physical parameters of a planetary gear set. **Figure 9** illustrates the mesh stiffness of one sun-planet gear pair and one ring-planet gear pair of this planetary gear set using the method given in Ref. [43]. Two health conditions are given: perfect condition and 4.3 mm tooth crack condition. This crack length is illustrated in **Figure 7**. The tooth crack is in the planet gear (ring side). Under this scenario, the mesh stiffness of the sun-planet gear pair is assumed not to be affected by this planet gear tooth crack [43], while the mesh stiffness of the ring-planet gear pair reduces.

Parameters	Sun	Planets (4, equally spaced)	Ring
Number of teeth	19	31	81
Module (mm)	3.2	3.2	3.2
Pressure angle	20°	20°	20°
Mass (kg)	0.700	1.822	5.982
Face width (m)	0.0381	0.0381	0.0381
Young's modulus (Pa)	2.068×10^{11}	2.068×10^{11}	2.068×10^{11}
Poisson's ratio	0.3	0.3	0.3
Base circle radius (mm)	28.3	46.2	120.8
Root circle radius (mm)	26.2	45.2	132.6
Reduction ratio	5.263		
Bearing stiffness	$k_{sx} = k_{sy} = k_{rx} = k_{ry} = k_{cx} = k_{cy} = k_{pmx} = k_{pmy} = 1.0 \times 10^{8}$ N/m		
Bearing damping	$c_{sx} = c_{sy} = c_{rx} = c_{ry} = c_{cx} = c_{cy} = c_{pmx} = c_{pmy} = 1.5 \times 10^{3}$ kg/s		

Table 2. Physical parameters of a planetary gear set [30].

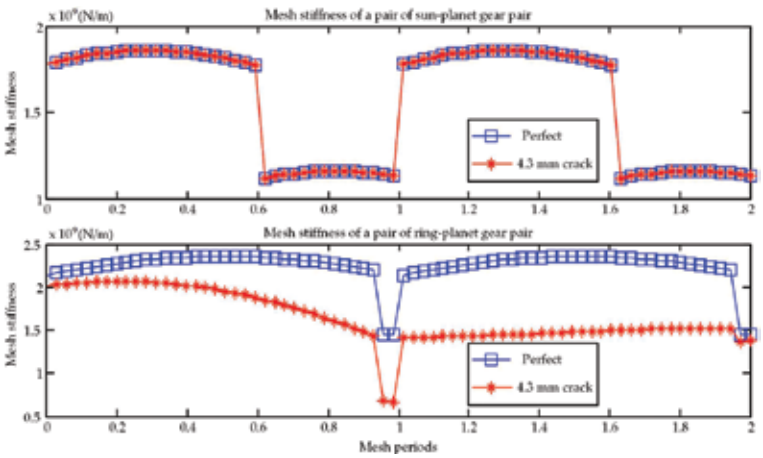


Figure 9. Mesh stiffness in perfect and cracked tooth conditions.

Similar methodology can be used to evaluate the effect of other gear faults like tooth pitting [36] on the mesh stiffness of a pair or gears. While many studies have been performed to evaluate gear mesh stiffness, the studies on the evaluation of gear mesh damping, bearing stiffness, and damping are rare. Gear mesh stiffness is assumed to be constant or proportional to gear mesh stiffness [40]. Bearing stiffness and damping are mostly assumed to be constant [30].

4. Modeling of transmission path effect

In **Figure 4**, three transmission paths are illustrated. However, most researchers only considered and modeled the effect of transmission path 1 since it has a shortest distance between vibration sources and the transducer. All researchers assumed that with the rotation of the carrier, the influence of a planet on vibration signals perceived by a transducer mounted on the gearbox housing reached its maximum when this planet was closest to the transducer; then this planet's influence decreased as the planet went away from the transducer. The transmission path effect model is assumed to be independent of gear fault modes. It can be used for gear tooth crack, pitting, spalling, wear, and so on. A Hanning function was used in Refs. [28, 29], and a Hamming function was used in Refs. [32, 50] to represent the effect of transmission path. A modified Hamming function with adjustable window bandwidth was proposed in Ref. [30]. Liu et al. [33] used the modified Hamming function to model the transmission path along the casing and also proposed two constants to represent the transmission path inside the casing. In this study, the modified Hamming function reported in Ref. [30] is used to represent the effect of transmission path. The resultant vibration $a(t)$ is modeled as the summation of weighted vibration of each planet gear as follows:

$$a(t) = \sum_{n=1}^{N} e^{a\left(\mathrm{mod}\left(w_c t+\phi,2\pi\right)-\pi\right)^2} H_m(t)\, a_n(t) \tag{2}$$

where N represents the number of planet gears, w_c denotes carrier rotation speed, ϕ represents circumferential angle of the nth planet gear, $H_m(t)$ is the Hamming function, $a_n(t)$ represents the acceleration signal of the nth planet gear, and the parameter a is used to control bandwidth of a Hamming function.

5. Vibration signal analysis and fault symptoms

Figure 10 gives the vertical acceleration signals of a single planet gear under healthy and cracked tooth conditions. The large spikes on this figure are generated by the meshing of the cracked tooth. The angular interval of these large spikes is 31 θ_m where θ_m represents the rotation angle of a planet gear in one tooth mesh period. The ring gear has 81 teeth, and therefore, in one revolution of the carrier, a planet gear's angular displacement is 81 θ_m.

Figure 11 shows the resultant vibration signals of a planet gearbox obtained using Eq. (2). Amplitude modulation can be observed from resultant vibration signals because four planet

http://dx.doi.org/10.5772/67529

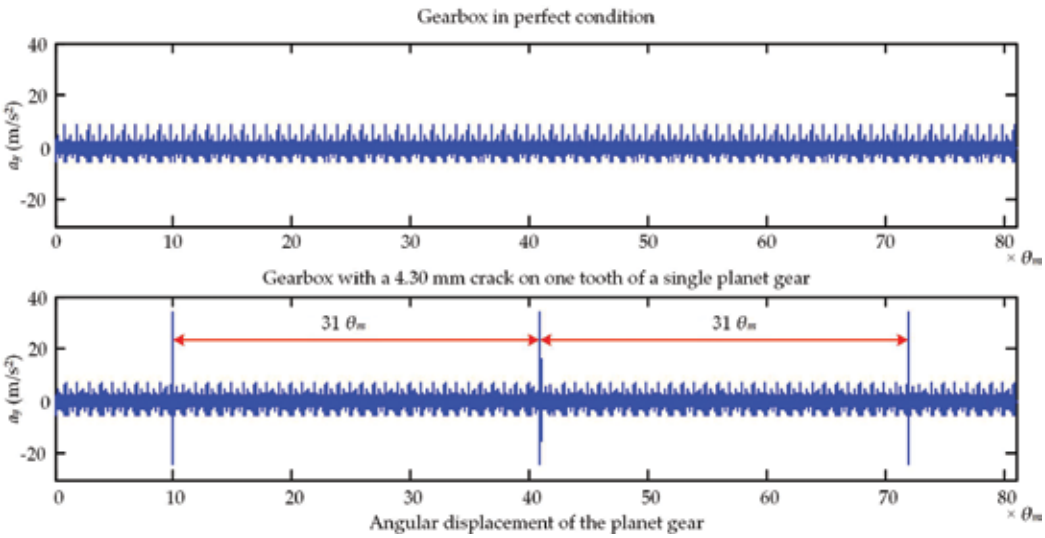


Figure 10. Vibration of a single planet gear.

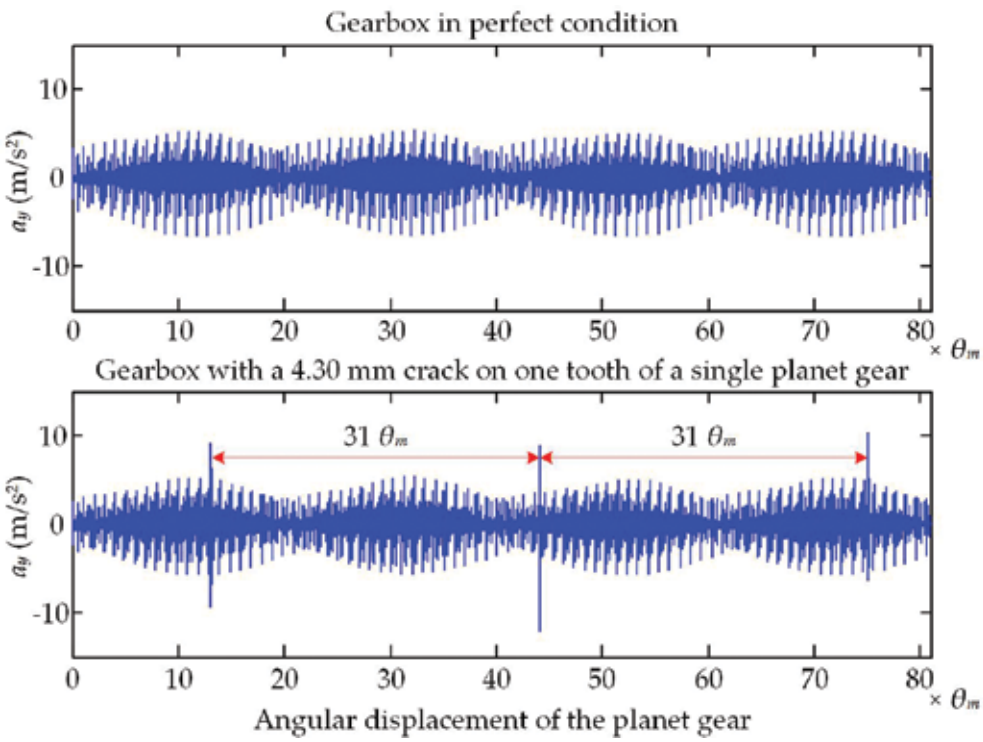


Figure 11. Resultant vibration of a planetary gearbox [31].

gears pass through the transducer sequentially in one revolution of the carrier and their effects on resultant vibration signals vary according to their locations. The amplitude of resultant vibration signals is smaller than that of a single planet gear because some vibration signals are attenuated during transmission. The amplitude of large spikes generated by the cracked tooth varies largely because the signal attenuation is not uniform. The farther between a planet gear and a transducer, the larger signal attenuation for this planet gear.

Similarly, for other types of planetary gear faults, the resultant vibration signals can be simulated by considering both vibration source signals and the effect of transmission path. The vibration source signals are related to gear faults by a dynamic model. The transmission path effect model is irrelevant with gear faults.

6. Vibration signal decomposition for fault diagnosis

From **Figure 11**, we can see that the time duration of large spikes generated by the cracked tooth is very short. These large spikes are not such obvious in experimental signals because of noise pollution [31]. Actually, they are very weak according to the experimental findings in Ref. [30, 31]. Therefore, people proposed signal decomposition techniques to enlarge the fault symptoms generated by a damaged gear tooth. In Ref. [51], McFadden used a window function to sample the vibration signals of a planetary gearbox when the planet gear of interest was passing the transducer and then the signal samples were mapped to the corresponding teeth of the sun gear or a planet gear to construct the vibration signals of the sun gear or a planet gear. The decomposed signal reduces the interference from other gear vibrations and emphasizes the fault symptoms of the gear of interest. Liang et al. [31] proposed another windowing and mapping strategy to generate the vibration signal of each tooth of the planet gear of interest. If a planet gear has N teeth, the resultant vibration signal can be decomposed into N subsignals. Each subsignal corresponds to one tooth. Examining the differences of these N subsignals, the health differences of these N teeth can be measured. The detailed signal decomposition techniques will not be described in detail in this chapter as they are available in the published papers [31, 51]. **Figure 12** illustrates the effectiveness of the windowing and mapping strategy proposed in Ref. [31]. **Figure 12(a)** gives the resultant vibration signal of a planetary gearbox with a single tooth crack on a planet gear. **Figure 12(b)** and **(c)** illustrate the decomposed vibration signals according to a perfect tooth and the cracked tooth, respectively. More obvious fault symptoms can be observed from **Figure 12(c)** than the original resultant vibration signal as shown in **Figure 12(a)**. Examining the differences between **Figure 12(b)** and **(c)**, it is easy to tell that **Figure 12(c)** is generated by a cracked tooth.

This vibration signal decomposition technique [31] is able to identify the tooth health condition differences. If some teeth are in healthy condition while others are in damaged condition, this method will be effective for fault diagnosis. In an extreme case, if all gear teeth have the same damage severity, like evenly distributed pitting, this method will be ineffective as there is no difference between the health conditions of teeth.

http://dx.doi.org/10.5772/67529

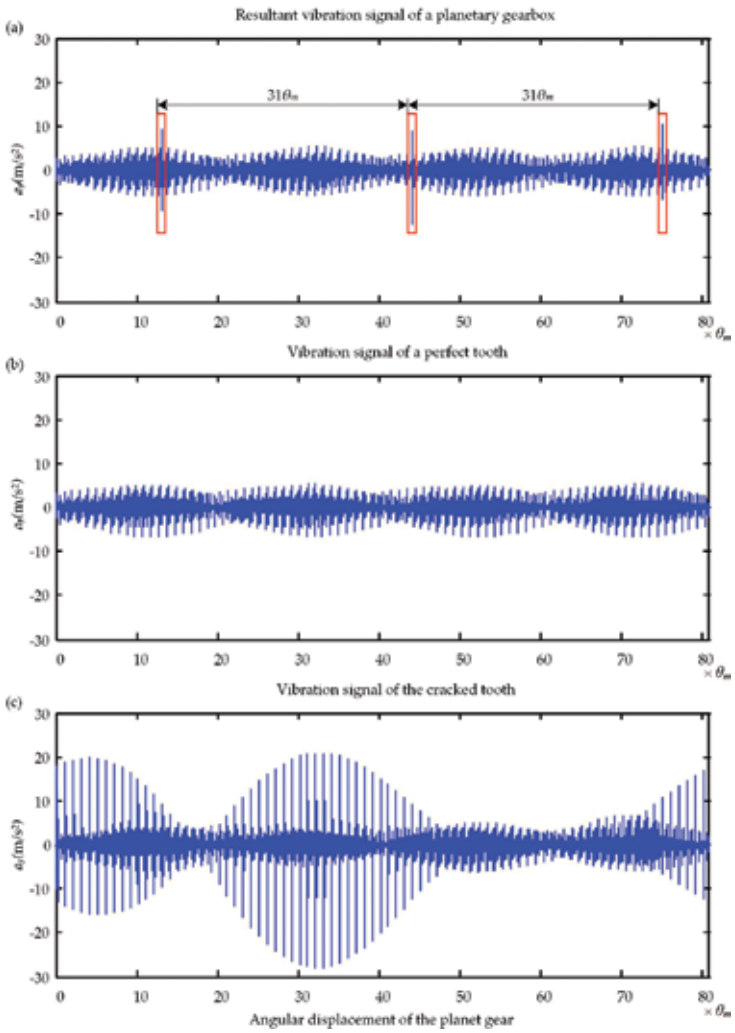


Figure 12. Vibration signal decomposition of a planetary gearbox [31]: (a) resultant vibration signal, (b) decomposed signal for a perfect tooth, and (c) decomposed signal for a cracked tooth.

7. Summary and future work

This chapter describes techniques for dynamics-based vibration signal modeling and fault diagnosis of planetary gearboxes. The planetary gearbox vibration signal modeling contains two parts: the simulation of vibration source signals and the modeling of transmission path effect. The current research status and challenges of dynamic modeling-based fault diagnosis are introduced. In the example given in this chapter, vibration source signals are obtained by a lumped parameter dynamic model, the gear fault is reflected by the time-varying mesh stiffness which is evaluated using the potential energy method, and the effect of transmission path is modeled using a modified Hamming function. Other window functions are also

described to model the effect of transmission path. However, further researches to test and validate these window functions are required. In the end, vibration decomposition techniques are briefly described for gear tooth damage diagnosis. Comparing between the raw signal and the decomposed signal, we can find that the signal decomposition technique can enlarge gearbox fault symptoms and facilitate fault diagnosis.

Based on the research scope of this chapter, the following perspectives are suggested for future consideration:

(1) Mesh stiffness evaluation with crack/pitting in multiple teeth.

(2) Mesh stiffness evaluation with multiple fault modes.

(3) Experimental validation of methods in evaluating gear mesh stiffness.

(4) Gear mesh damping evaluation.

(5) Bearing stiffness and damping evaluation.

(6) Time-varying load or random load effect on the vibration signals.

(7) Experimental validation of models for transmission path effects.

(8) Effects of noises from internal and/or external sources.

(9) Dynamic and vibration properties of gearboxes with multiple faults (multiple fault locations and/or multiple fault modes).

(10) Development of fault diagnosis techniques based on the understanding of vibration properties.

Acknowledgements

This research is supported by the Natural Science and Engineering Research Council of Canada, Canada (Grant No. RGPIN-2015-04897); the International S&T Cooperation Program of China, China (Grant No. 2015DFA71400); the National Natural Science Foundation of China, China (Grant No. 51375078 and No. 51505066); the ConocoPhillips Canada Limited Graduate Scholarship, Canada; and the American Gear Manufacturers Association Foundation Scholarship, USA.

Author details

Xihui Liang[1], Ming J. Zuo[1]* and Wenhua Chen[2]

*Address all correspondence to: ming.zuo@ualberta.ca

1 Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada

2 Key Laboratory of Reliability Technology for Mechanical and Electrical Product of Zhejiang Province, Zhejiang Sci-Tech University, Hangzhou, China

http://dx.doi.org/10.5772/67529

References

[1] The Guardian. North Sea helicopter crash report says gearbox failed after maintenance error [Internet]. 2011. Available from: https://www.theguardian.com/uk/2011/nov/24/north-sea-helicopter-crash-report [Accessed: 2017-02-14]

[2] Hoseini MR, Lei Y, Tuan DV, Patel T, Zuo MJ. *Experiment design of four types of experiments: pitting experiments, run-to-failure experiments, various load and speed experiments, and crack experiments*. Technical Report, Reliability Research Lab, Mechanical Department, University of Alberta, Edmonton, Alberta, Canada, January 2011.

[3] Zhao X, Zuo MJ, Liu Z, Hoseini MR. Diagnosis of artificially created surface damage levels of planet gear teeth using ordinal ranking. *Measurement* 2013; 46: 132–144.

[4] Liu Z, Qu J, Zuo MJ, Xu H. Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel fisher discriminant analysis. *Int J Adv Manuf Technol* 2012; 67: 1217–1230.

[5] Feng Z, Zuo MJ. Vibration signal models for fault diagnosis of planetary gearboxes. *J Sound Vib* 2012; 331: 4919–4939.

[6] Feng Z, Zuo MJ. Fault diagnosis of planetary gearboxes via torsional vibration signal analysis. *Mech Syst Signal Process* 2013; 36: 401–421.

[7] Lin J, Parker RG. Analytical characterization of the unique properties of planetary gear free vibration. *J Vib Acoust* 1999; 121: 316–321.

[8] Ambarisha VK, Parker RG. Nonlinear dynamics of planetary gears using analytical and finite element models. *J Sound Vib* 2007; 302: 577–595.

[9] Bahk C-J, Parker RG. Analytical solution for the nonlinear dynamics of planetary gears. *J Comput Nonlinear Dyn* 2011; 6: 21007.

[10] Parker RG. A physical explanation for the effectiveness of planet phasing to suppress planetary gear vibration. *J Sound Vib* 2000; 236: 561–573.

[11] Ericson T, Parker RG. Planetary gear modal properties and dynamic response: experiments and analytical simulation. In: *ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, Columbus, Ohio, USA, 2011. p. 331–343.

[12] Ericson TM, Parker RG. Planetary gear modal vibration experiments and correlation against lumped-parameter and finite element models. *J Sound Vib* 2013; 332: 2350–2375.

[13] Al-shyyab A, Kahraman A. A non-linear dynamic model for planetary gear sets. *Proc Inst Mech Eng Part K J Multi-Body Dyn* 2007; 221: 567–576.

[14] Inalpolat M, Kahraman A. Dynamic modelling of planetary gears of automatic transmissions. *Proc Inst Mech Eng Part K J Multi-Body Dyn* 2008; 222: 229–242.

[15] Yuksel C, Kahraman A. Dynamic tooth loads of planetary gear sets having tooth profile wear. *Mech Mach Theory* 2004; 39: 695–715.

[16] Kahraman A. Load sharing characteristics of planetary transmissions. *Mech Mach Theory* 1994; 29: 1151–1165.

[17] Kahraman A Planetary gear train dynamics. *J Mech Des* 1994; 116: 713–720.

[18] Chen Z, Shao Y. Dynamic features of a planetary gear system with tooth crack under different sizes and inclination angles. *J Vib Acoust* 2013; 135: 1–12.

[19] Shao Y, Chen Z. Dynamic features of planetary gear set with tooth plastic inclination deformation due to tooth root crack. *Nonlinear Dyn* 2013; 74: 1253–1266.

[20] Chen Z, Shao Y. Dynamic simulation of planetary gear with tooth root crack in ring gear. *Eng Fail Anal* 2013; 31: 8–18.

[21] Cheng Z, Hu N, Zhang X. Crack level estimation approach for planetary gearbox based on simulation signal and GRA. *J Sound Vib* 2012; 331: 5853–5863.

[22] Cao Z, Shao Y, Zuo MJ, Liang X. Dynamic and quasi-static modeling of planetary gear set considering carrier misalignment error and varying line of action along tooth width. *Proc Inst Mech Eng Part C J Mech Eng Sci* 2015; 229: 1348–1360.

[23] Cheng Z, Hu N, Gu F, Qin G. Pitting damage levels estimation for planetary gear sets based on model simulation and grey relational analysis. *Trans Can Soc Mech Eng* 2011; 35: 403–417.

[24] Li G, Li F, Wang Y, Dong D. Fault diagnosis for a multistage planetary gear set using model-based simulation and experimental investigation. *Shock Vib* 2016; 2016: 1–19.

[25] Chaari F, Fakhfakh T, Haddar M. Dynamic analysis of a planetary gear failure caused by tooth pitting and cracking. *J Fail Anal Prev* 2006; 6: 73–78.

[26] Chaari F, Hbaieb R, Fakhfakh T, Haddar M. Dynamic response simulation of planetary gears by the iterative spectral method. *Int J Simul Model* 2005; 4: 35–45.

[27] Chaari F, Fakhfakh T, Hbaieb R, Louati J, Haddar M. Influence of manufacturing errors on the dynamic behavior of planetary gears. *Int J Adv Manuf Technol* 2006; 27: 738–746.

[28] Inalpolat M, Kahraman A. A theoretical and experimental investigation of modulation sidebands of planetary gear sets. *J Sound Vib* 2009; 323: 677–696.

[29] Inalpolat M, Kahraman A. A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors. *J Sound Vib* 2010; 329: 371–393.

[30] Liang X, Zuo MJ, Hoseini MR. Vibration signal modeling of a planetary gear set for tooth crack detection. *Eng Fail Anal* 2015; 48: 185–200.

[31] Liang X, Zuo MJ, Liu L. A windowing and mapping strategy for gear tooth fault detection of a planetary gearbox. *Mech Syst Signal Process* 2016; 80: 445–459.

[32] Liang X, Zuo MJ. *Investigating Vibration Properties of a Planetary Gear Set with a Cracked Tooth in a Planet Gear*. Fort Worth, Texas, USA, 2014. p. 568–575.

[33] Liu L, Liang X, Zuo MJ. Vibration signal modeling of a planetary gear set with transmission path effect analysis. *Measurement* 2016; 85: 20–31.

[34] Ritger PD, Rose NJ. *Differential equations with applications*. Courier Corporation, USA 1968.

[35] Parker RG, Vijaykar SM, Imajo T. Non-linear dynamic response of a spur gear pair: modelling and experimental comparisons. *J Sound Vib* 2000; 237: 435–455.

[36] Liang X, Zhang H, Liu L, Zuo MJ. The influence of tooth pitting on the mesh stiffness of a pair of external spur gears. *Mech Mach Theory* 2016; 106: 1–15.

[37] Yang DCH, Lin JY. Hertzian damping, tooth friction and bending elasticity in gear impact dynamics. *J Mech Des* 1987; 109: 189–196.

[38] Liang X, Zuo MJ, Patel TH. Evaluating the time-varying mesh stiffness of a planetary gear set using the potential energy method. *Proc Inst Mech Eng Part C J Mech Eng Sci* 2014; 228: 535–547.

[39] Tian X, Zuo MJ, Fyfe KR. Analysis of the Vibration Response of a Gearbox with Gear Tooth Faults. Anaheim, California, USA, 2004. p. 785–793.

[40] Wu S, Zuo MJ, Parey A. Simulation of spur gear dynamics and estimation of fault growth. *J Sound Vib* 2008; 317: 608–624.

[41] Chen Z, Shao Y. Dynamic simulation of spur gear with tooth root crack propagating along tooth width and crack depth. *Eng Fail Anal* 2011; 18: 2149–2164.

[42] Pandya Y, Parey A. Failure path based modified gear mesh stiffness for spur gear pair with tooth root crack. *Eng Fail Anal* 2013; 27: 286–296.

[43] Liang X, Zuo MJ, Pandey M. Analytically evaluating the influence of crack on the mesh stiffness of a planetary gear set. *Mech Mach Theory* 2014; 76: 20–38.

[44] Ma H, Song R, Pang X, et al. Time-varying mesh stiffness calculation of cracked spur gears. *Eng Fail Anal* 2014; 44: 179–194.

[45] Ma H, Pang X, Zeng J, Wang Q, Wen B. Effects of gear crack propagation paths on vibration responses of the perforated gear system. *Mech Syst Signal Process* 2015; 62–63: 113–128.

[46] Abouel-seoud SA, Dyab ES, Elmorsy MS. Influence of tooth pitting and cracking on gear meshing stiffness and dynamic response of wind turbine gearbox. *Int J Sci Adv Technol* 2012; 2: 151–165.

[47] Ma H, Li Z, Feng M, Feng R, Wen B. Time-varying mesh stiffness calculation of spur gears with spalling defect. *Eng Fail Anal* 2016;66: 166–176.

[48] Chen Z, Shao Y. Mesh stiffness calculation of a spur gear pair with tooth profile modification and tooth root crack. *Mech Mach Theory* 2013;62: 63–74.

[49] Parker RG, Lin J. Mesh phasing relationships in planetary and epicyclic gears. *J Mech Des* 2004; 126: 365–370.

[50] Liang X, Zuo MJ, Hoseini MR. Understanding vibration properties of a planetary gear set for fault detection. In: *2014 IEEE Conference on Prognostics and Health Management (PHM)*. 2014, pp. 1–6.

[51] McFadden PD. A technique for calculating the time domain averages of the vibration of the individual planet gears and the sun gear in an epicyclic gearbox. *J Sound Vib* 1991; 144: 163–172.

Chapter 8

Fault-Tolerant Electrical Machines and Drives

Mircea Ruba

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67354

Abstract

The last years of research and development in the automotive industry were still focused on designing electrical propulsion units to be eco-friendly and diminish the drawbacks of classical combustion engines. Besides being energy efficient, silent, and high in power density, these must have a serious fault-tolerant ability as driver, and passengers' safety is probably the most important issue in this filed. The chapter will detail fault-tolerant machines and power electronic architectures with their control for the most common ones, such as switched reluctance machines (SRM) and the permanent-magnet synchronous machines (PMSM). Besides detection, solutions will be presented for the machine-drive unit to wisely overcome and compensate occurred faults. A novel modular structure of SRM is presented with increased fault tolerance and possibility of fast repair in case of any machine damage. The solutions will be validated via simulated and experiment-based results.

Keywords: fault tolerant, electrical machines, drives, control, fault detection, fault compensation, SRM, PMSM

1. Introduction

Electromechanical systems became in the last decade an indispensable part of nearly every electrical energy consumer. Taking advantage of the technological advance in the world of engineering, the demands in terms of performance, efficiency, reliability, and safety which were raised are continuously pushed forward [1]. Latest technological developments for such systems are based on sophisticated control strategies that have the ability to accommodate component failures automatically. By this, the system's stability is maintained in acceptable range of performances reaching what one can entitle fault-tolerant system [2].

This chapter focuses on switched reluctance and permanent-magnet synchronous machines (SRM and PMSM), dedicated for fault-tolerant applications. Modular and multiphase designs will be presented for both types of machines together with their electronic converters,

developed to serve such applications. The motivation for choosing these two electrical machines and their drives comes with the tendency of the actual market to focus research to electrify the propulsion systems in the automotive industry. Researches involved complex studies in developing solutions for fault-tolerant SRMs with their drives, solutions that are flexible and able to do online fault diagnosis and compensation [3]. Other studies were focused on electronic failure diagnosis, like the one in Ref. [4]. Winding malfunctions, such as short circuits [5] and open circuits [6], were engaged, offering solutions for intelligent control strategies. The proposed modular SRM adds to the actual status of research by a novel structural design that allows isolation and easy replacement of the faulted coils.

In the field of PMSMs, winding fault is usually the main cause of malfunctions due to short or open circuit faults. Usually, these are observed and compensated by implementing intelligent control strategies [7, 8]. Research activities were carried out also in the field of electronic faults, such as short circuit, proposing control strategies to avoid the error and try to keep the machine operational [9]. Field Programmable Gate Array (FPGA) based control methods for fault recognition and compensation based on a residual calculation were proposed in Ref. [10]. The multiphase PMSM detailed in the chapter offers a simple motor-drive fault-tolerant solution that by its architecture comes in addition to the preliminary studies.

2. Fault-tolerant switched reluctance machines and drives

By the nature of operation and by their construction, SRMs are considered up to an extent fault tolerant because these are able to operate in case of one-phase failure. However, adding to its natural fault-tolerant ability a dedicated modular stator design, one can reach a high level of safety and reliability when the application demands it. Even more, adding to such a machine a more complex electronic power converter, able to isolate and compensate the occurred fault, increases radically the fault-tolerant abilities.

The modular design of the SRM's stator consists of eight individual modules isolated between them with nonmagnetic displacers as shown in **Figure 1**. One machine phase is compounded of the coils wound on diametrically opposed modules. These are fixed to the end caps of the machine using nonmagnetic rods that pass through them. Using the nonmagnetic displacers between adjacent modules, the magnetic flux's path is forced only via the energized module and the corresponding rotor poles as seen in **Figure 1**. The machine's rotor is a classical design having 14 poles.

The concept regards that the coils of each phase are fed independently. In this case, with independent supply and because one stator slot is not shared by coils from different phases, in case of short circuits, overheating, etc. of one coil, the remaining ones are not influenced nor altered.

This way, if a coil of one phase is faulted, the machine will continue operating with the remaining diametrically opposed coil. This event can occur from one to all phases of the machine, and it can still operate in satisfactory conditions. One solution to compensate the faulted coils, is to increase the current of the remaining healthy ones. This is allowed by

http://dx.doi.org/10.5772/67354

their design, even if it will mean saturation of the core and increase of the machine's temperature. It is easy to understand that the developed torque and speed in case of fault will be lower than in normal condition but, still, will reach satisfactory values till safe stop of the machine is possible.

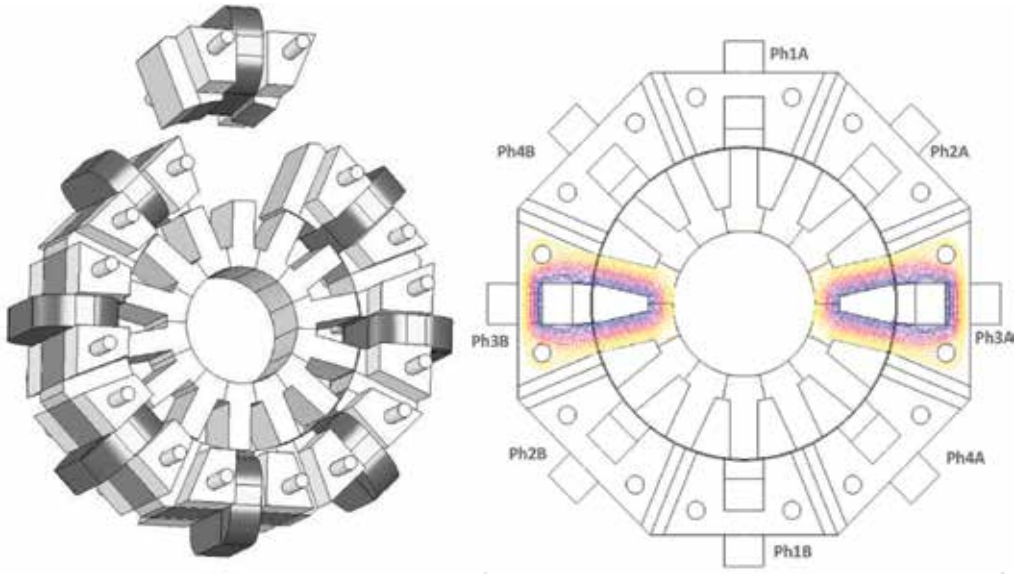


Figure 1. The 3D model of the modular SRM and its flux paths on one phase and the coil distribution.

This design allows besides isolation of the coils and of the magnetic cores simple and fast replacement of the damaged modules with new ones. Extracting the fixing rods permits removal of the nonfunctional modules without the need of uncoupling the shaft from the application and without moving the machine out of its housing. The prototype built to prove the concept was rated at 300 Vdc, 6 A, 600 rpm, and 5 Nm, as specifications for the design process.

2.1. The design of the modular SRM

Due to the fact that the machine has nonclassical features, its design process is based on an analytical model combined with particular calculations required by the modular structure. The number of rotor poles (Q_R) is a function of the number of stator modules (N_{mS}) and the number of phase to coil division (n_{div}):

$$Q_R = n_{div} \cdot N_{mS} - n_{div} \tag{1}$$

The design process followed complex algorithms [11–13] to shape the geometry of the core assemblies but at the same time, to size and determine the machine's characteristics such as torque, losses, flux densities, etc.

The shaping process starts by imposing several parameters that will characterize the performances of the future machine. These regard the supply voltage (U_N), the rated current (I), the number of phases (m), the machine's rated power (P_{2N}), the mechanical air gap (g), the air-gap flux density in aligned position (B_{gmax}), the rated speed (n_N), and the rated torque (T_N).

For any SRM, classical or modular one, the most important parameter that has important influence on its performances is the mean diameter, as value measured in the middle point of the air gap [14]:

$$D_g = \sqrt[3]{\frac{P_{2N} \cdot Q_s \cdot k_\sigma}{Q_R \cdot \pi^2 \cdot k_L \cdot \frac{n_N}{60} \cdot B_{g\max} \cdot \left(1 - \frac{1}{K_\sigma}\right) \cdot A_S}} \tag{2}$$

where (Q_S) and (Q_R) are the numbers of stator and rotor poles. Coefficients (k_σ) and (k_L) regard the leakage flux factors, as values between 0.75 and 0.95, the aspect factor, respectively, are calculated like

$$k_L = \frac{\pi}{2} \cdot \frac{1}{\sqrt[3]{Q_R}} \tag{3}$$

The electrical loading A_S will be chosen in an interval between 5^4 and 15^4 A/m, while Carter's factor K_{CR} considers the shape of the salient poles and the flux path's distortion due to this shape [15]. It is chosen in a range between 1.4 and 2.

Computing the ratio of the mean diameter (D_g) with respect to the aspect of coefficient yells for a preliminary value of the machine's active stack length (l_a) that can be later adjusted as function of the resulted developed torque. Usually, this value can be restricted by the housing of the machine.

Because the stator is compounded of eight independent modules, it is enough to size only one of them. The complete circle described by the inner stator diameter is split into (N_{mS}) arcs, equal with the number of the stator modules. The length of one arch described by one module is

$$L_m = \frac{2 \cdot \pi \cdot \left(\frac{D_g}{2} + \frac{g}{2}\right)}{N_{mS}} \tag{4}$$

The SRM by its nature due to its saliency and due to its switched current operation has the well-known torque ripples, influenced also by the saturation of the pole tips. When discussing about analytical calculation of the mean torque, it is quite difficult to take into account all the factors. Hence, as the most influential parameter on the torque development is the core saturation, instead of the geometrically defined air gap (g), an equivalent value will be used in the following calculations, the latter considering the saturation of the poles in aligned position by (k_{sat}) coefficient:

$$g_x = k_{sat} \cdot g \tag{5}$$

To ensure correct isolation between the modules, it was considered necessary that the spacers between them (l_D) should be at least 10 but not larger than 20 times the air gap.

http://dx.doi.org/10.5772/67354

The arch defined by one stator module (see **Figure 2**) computed with Eq. (4) includes also the arch defined by the spacer. Hence, the pole pitch described just by the module itself without the spacer is half of the difference between (L_m) and (l_D):

$$\tau_S = \frac{L_m - l_D}{2} \tag{6}$$

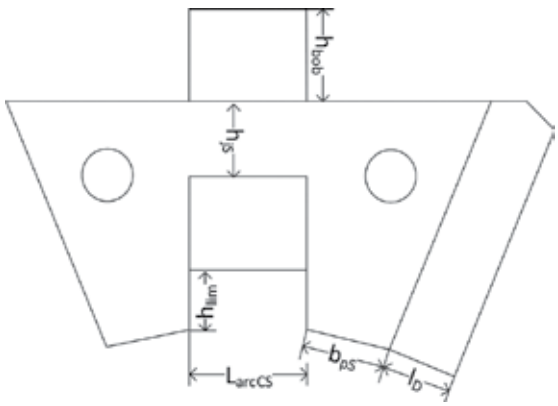


Figure 2. Sizing the stator module.

Sizing the width of the stator poles is based on ranging their values between 0.5 and 0.8 of the module total pitch. The value has to be considered in order to leave space to fit the winding and to ensure correct overlapping with the rotor poles. Round values are indicated to be used, as it simplifies the building process. Hence, the stator pole width was considered to be

$$b_{pS} = round(0.58 \cdot \tau_S) \tag{7}$$

Finalizing the computation of the stator pole width requires to recompute the value of the spacers between the modules, in order to close the circle described by the inner stator radius. By now, only a preliminary value of the spacer was considered. Before proceeding to this calculation, firstly the rotor pole pitch has to be computed:

$$L_{arcPR} = \frac{\pi \cdot \left(\frac{D_g}{2} - \frac{g}{2}\right) \cdot \frac{360}{Q_R}}{180} - 2 \cdot \frac{b_{pR}}{2} \tag{8}$$

where b_{pR} is the rotor pole width, considered equal with the stator pole width (b_{pS}). To calculate the real dimensions of the spacer, the stator slot opening width (L_{arcCS}) needs to be computed:

$$L_{arcCS} = round\left(\frac{\pi\left(\frac{D_g}{2} + \frac{g}{2}\right) \cdot u_C}{180}\right) \tag{9}$$

where (u_C) is the angle described by the rotor pole pitch arc, calculated as

$$u = \frac{\pi \cdot \left(\frac{D_g}{2} - \frac{g}{2}\right) - \frac{360}{Q_R} - b_{pS}}{\pi \cdot \left(\frac{D_g}{2} - \frac{g}{2}\right)} \tag{10}$$

To finalize the sizing of the module spacers, it is considered that the stator and rotor slots are equal and the distance between them is just the air gap. Hence, the spacer can be sized as

$$l_D = L_m - 2 \cdot b_{pS} - L_{arcCS} \tag{11}$$

By now, only the module's yoke height remains to be sized. To make sure that proper saturation is reached advantaging fast demagnetization when the phase is switched in *off* state, the yoke was considered to be 0.85 of the pole's width:

$$h_{jS} = round\left(0.85 \cdot b_{pS}\right) \tag{12}$$

Because no information is yet available about the size of the coils that need to be fitted inside the stator slot, only a preliminary value can be considered to the height of the stator poles, simply obtained as

$$h_{pS} = round(1.01 \cdot L_m) \tag{13}$$

This value will be recomputed after sizing the coils. However, the active stator pole surface will be

$$A_{pS} = b_{pS} \cdot l_a \tag{14}$$

Before sizing the coils, which will require information about the length of the flux paths, the rotor must be sized. As already stated, the rotor pole pitch is equal with the one of the stator. As the width of the rotor poles are also at the same size as the one of the stator poles, one can compute the rotor slot opening like

$$b_{cR} = round\left(L_{arcR} - b_{pR}\right) \tag{15}$$

As seen in **Figure 1**, the magnetic flux closes only via the energized module and the corresponding rotor poles; hence, the yoke of the rotor (h_{jR}) can be considered equal with the one of the stator (h_{jS}). Imposing the diameter of the machine's shaft (d_{ax}), it is possible to compute rotor pole height (see **Figure 3**):

$$h_{pR} = round\left(\frac{D_g}{2} - \frac{g_x}{4} - h_{jR} - \frac{d_{ax}}{2}\right) \tag{16}$$

As mentioned earlier, after sizing the rotor, it is possible to compute the dimensions of the coils in relation with the magnetic field (H) and the machine's geometrical dimensions.

To be able to compute the required magnetomotive force, the length of the flux paths through the stator (l_s), air gap (l_g), and rotor (l_r) must be calculated:

http://dx.doi.org/10.5772/67354

$$\begin{aligned} l_s &= 2 \cdot h_{pS} + L_{arcCS} \\ l_g &= 2 \cdot g \\ l_r &= 2 \cdot h_{pR} + L_{arcCR} \end{aligned} \tag{17}$$

Based on Ampere's law, it is easy to compute the magnetomotive force, and from there the number of turns per coil N_f is

$$\begin{aligned} \Theta &= H_{Fe} \cdot (l_s + l_r) + H_g \cdot l_g \\ N_f &= round(\Theta / I) \end{aligned} \tag{18}$$

Function of the desired current and the number of turns, one can easily determine the size of the coil (h_{bob}) that will allow the final computation of the height of the stator module:

$$h_m = h_{\lim} + h_{bob} + h_{jS} \tag{19}$$

Having all the abovementioned data, sizing process of the machine is finished. To validate the breviary, the designer has to compute the developed mean torque using

$$T_v = k_{unal} \cdot N_{op} \left(N_f \cdot I\right)^2 \cdot \frac{D_g}{2} \cdot \mu_0 \cdot \frac{l_a}{4 \cdot g_x} \tag{20}$$

where (N_{op}) is the number of modules of one phase and (k_{unal}) is a constant that considers the contribution of the magnetic flux in unaligned position.

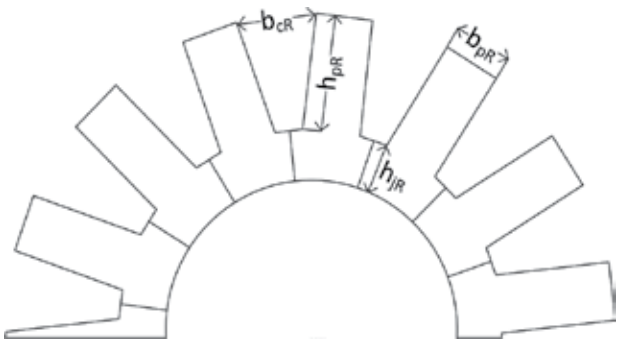


Figure 3. Sizing the rotor of the machine.

2.2. The operation of the fault-tolerant SRM

Validation of the modular SRM's design was accomplished via both finite element simulations and experimental testing in the laboratory. In both cases, the electronic converter had half-H bridges for each coil like in **Figure 4**. Hence, the two coils of one phase had each an independent half-H bridge (e.g., phase Ph1 is compounded of two independently but synchronously supplied coils, Ph1A and Ph1B). The concept was to permit the converter and the controller to separate and compensate the faulted coils. During normal operation, the coils of one phase were supplied with the same current using hysteresis current controller.

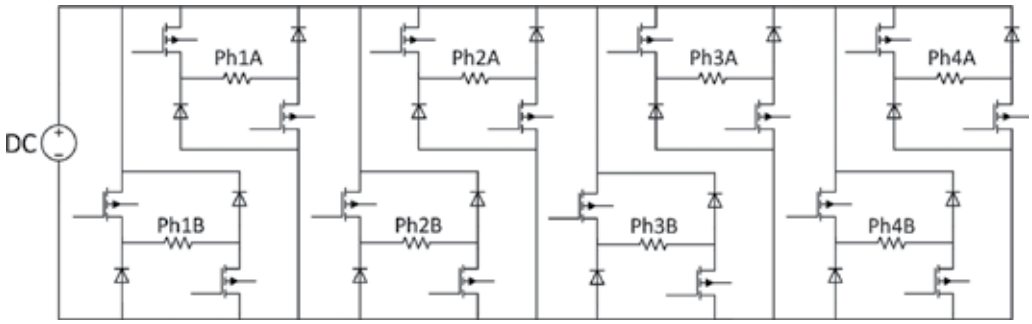


Figure 4. The electronic converter's topology.

It is clear that in case of any fault, the controller stops feeding with current the damaged coil and supplies only the healthy remaining one. This structure is indeed a more complex one than in the case of classical converters, but it is motivated by the highly increased fault-tolerant application of the machine.

Using Flux2D software, several tests were performed in both healthy and faulty conditions with 1–4 open circuit coils. In **Figure 5(a)**, **(b)**, and **(c)**, the healthy condition is depicted. The currents are plotted separately for all the eight coils of the machine (coils Ph1A, Ph2A, Ph3A, and Ph4A in plot b) and (coils Ph1B, Ph2B, Ph3B, and Ph4B in plot c). The developed RMS torque is 5.7 Nm at the rated current of 6 A. In case of one faulted coil (open circuit of Ph1A of phase 1), as it can be seen in **Figure 5d**, correspondence to the missing coil's (**Figure 5e**) current of the torque falls, being generated only by Ph1B of phase 1. Due to separate supply of the two coils of the phases (see **Figure 4**), Ph1B remains operational, as it is depicted in **Figure 5f**. The machine continues to operate even if the ripples are increased. It has to be mentioned also that all the tests were carried out only with classical hysteresis control strategy just to prove the machine's operational skills.

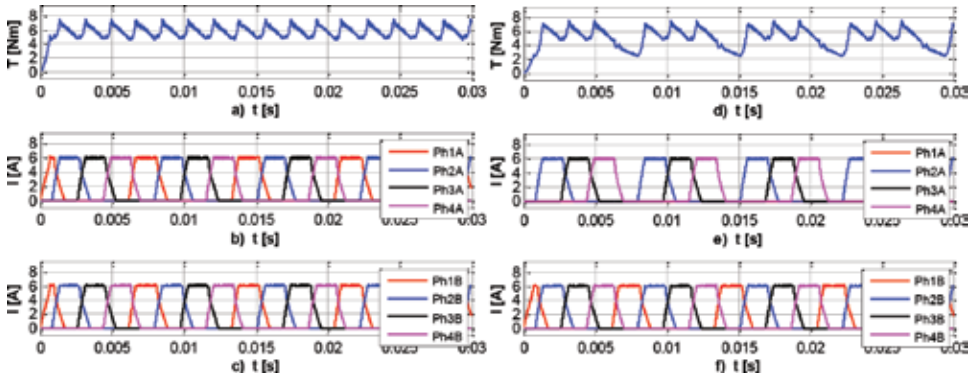


Figure 5. Finite element simulation results in healthy (a–c) and one-coil open fault (d–f).

In case of more sever faults, as it can be seen in **Figure 6** for two to four opened coils, the RMS torque decreases and yells for more ripples, still the machine being able to perform continuous operation.

http://dx.doi.org/10.5772/67354

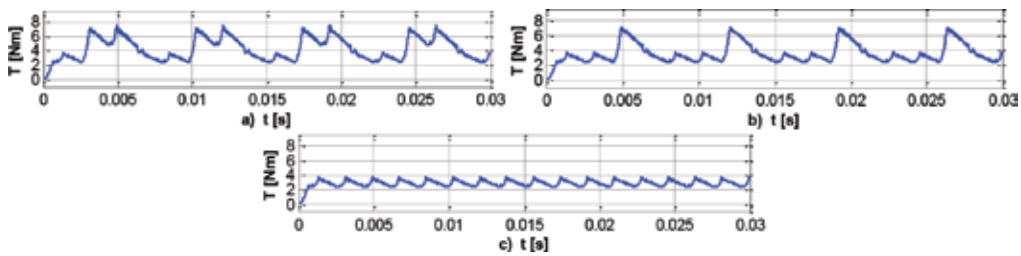


Figure 6. Simulated results in faulted conditions: two coils (a), three coils (b), and four coils (c).

The experimental validation of the concept and of the machine itself was accomplished by building the modular SRM with regard to the design specifications and setting up a test bench like the one in **Figure 7**. The modules were cut out of lamination and fixed into an aluminum housing. The test bench was compounded of the SRM, an induction machine as a load, a torque meter, the complex converter, and a dSPACE 1104 card for its control. An encoder was mounted on the shaft of the machine to provide feedback of the speed and the rotor position. To prove the operational skills of the machine, the same tests were carried out on the test bench like those performed in simulations.

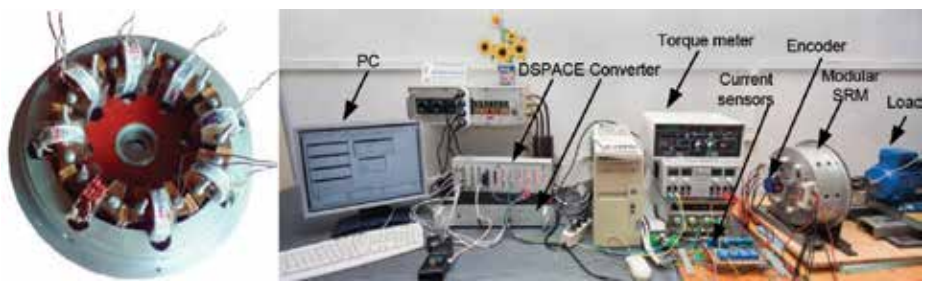


Figure 7. The modular stator of the SRM (left) and the experimental test bench (right) [8].

The experimental results depicted in **Figure 8(a)** to **(c)** in healthy condition and **Figure 8(d)** to **(f)** in open fault condition prove that the machine reaches the same performances as in simulations. As the measured results are quite close to the ones obtained in simulations, the same explanations that were detailed with regard to **Figure 5** are valid for the measured results.

In **Figure 9**, the results for faulted condition in experimental testing are depicted. The same remarks described for the simulated results of each fault scenario are valid also for the measured ones. By this, it was proved that the design breviary and the experimental model of the SRM are in accordance and the machine reaches the desired performances. In case of faults, if the currents on the remaining coils are not increased, the torque ripples are intensified. However, the machine can continue its operation even if four of the eight coils are not operational. The slight differences between the simulated results in **Figure 5** and those obtained in real measurements in **Figure 8** are due to low sampling speed and quite low precision in rotor position measurement, as an encoder was used instead of a resolver.

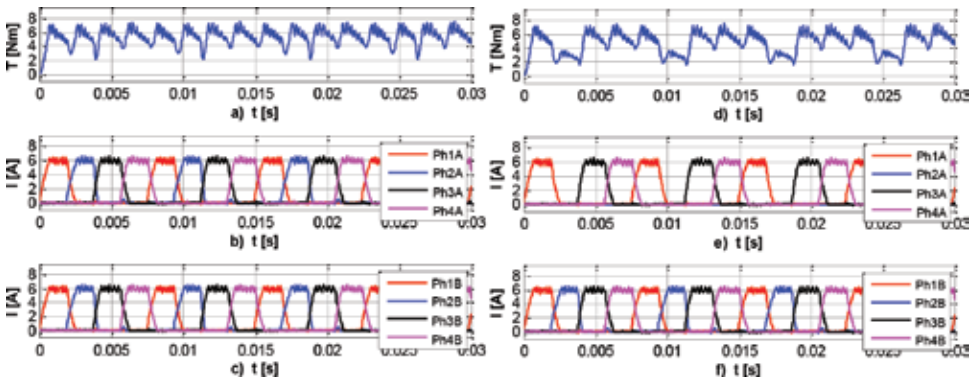


Figure 8. Experimental results in healthy (a–c) and one-coil open fault (d–f).

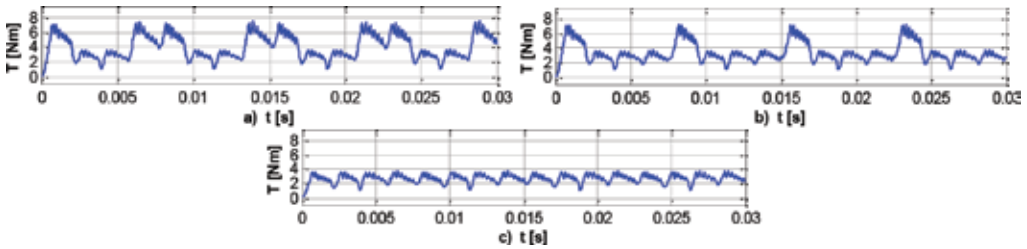


Figure 9. Experimental results in faulted conditions: two coils (a), three coils (b), and four coils (c).

Comparing the results obtained from simulations and those obtained in experimental testing, one can conclude that even in faulty condition, the machine reaches quite good performances. Comparing the results gathered in **Table 1**, it can be stated that there are quite good agreements between them. The biggest difference was reached in four opened coil conditions, about 0.65 Nm. However, in any fault condition, the SRM is able to develop more than 60% of the rated torque, this correspondence to the worst case scenario (four opened coils).

	RMS torques (N·m)	
Condition	**Simulated**	**Measured**
Healthy machine	5.5	5.3
One faulted coil	4.9	4.68
Two faulted coils	4.34	4.05
Three faulted coils	3.8	3.32
Four faulted coils	3.3	2.65

Table 1. The RMS torque obtained in simulated and experimental analysis.

http://dx.doi.org/10.5772/67354

Proving the concept and the operational skills, it is worth mentioning that, due to simple hysteresis control strategy, in case of faults, the torque ripples are quite high. However, engaging torque-smoothing strategies, such as direct instantaneous torque control or current profiling, it is possible to reach satisfactory torque characteristic, quite close to a linear one [9, 10], even in case of fault operation.

3. Fault-tolerant permanent-magnet synchronous machines and drives

The PMSMs due to their high power density, maturity, and reliability became in the last decade more and more used in all the fields of electromechanical systems. Many applications such as aircrafts, military, medicine, electric propulsion, etc. embed such machines, demanding in the same time high fault-tolerant capabilities. Hence, fault-tolerant PSMSs became a hot topic. Investigating ideas like six-phase inverters with common mode voltage elimination [16], or using series-connected six-phase inverter with a three-phase motor to use the same inverter for two machines [17] or adding a supplementary leg to the three-phase inverter, used only in faulty condition [18, 19] became widely used solutions for fault PMSM-tolerant applications. However smart-design considerations [20, 21] and dedicated control strategies [22, 23] were applied to increase the fault-tolerant potential of the machine.

In order to achieve a more increased level of fault tolerance, a nine-phase PMSM is proposed supplied from a special electronic converter able to overcome and compensate the occurred fault, like the one in **Figure 10**. The design of the machine, the power converter, and their operation will be described in the following, highlighting the benefits of using such a complex but fault-free system.

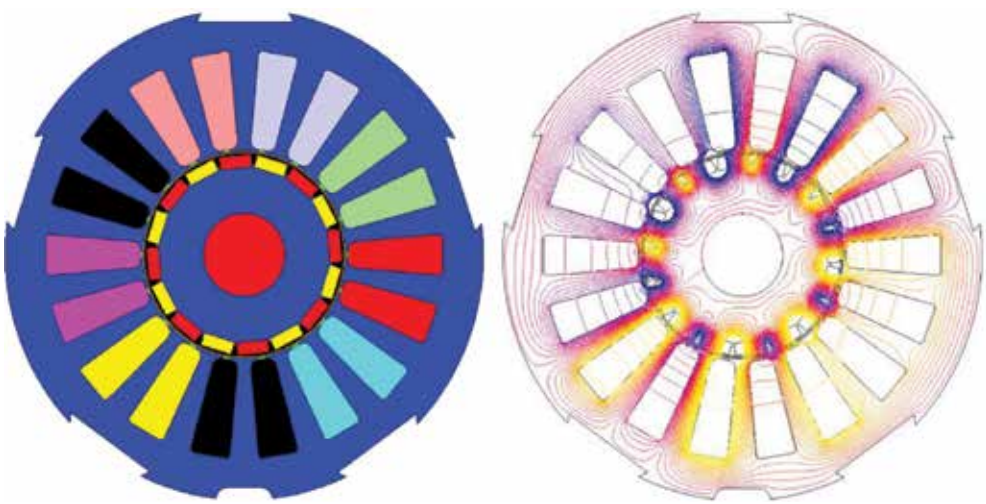


Figure 10. The structure of the nine-phase PSMSM (left) and its flux paths (right).

3.1. The design of the nine-phase PMSM

Taking into account published solutions to reduce significantly the torque ripples using fractional-slot winding configuration [24, 25], a PMSM with 8 pole pairs and 18 slots was designed. The machine has nine phases, and in the addition to classical designs, each phase is formed by one coil that surrounds only one tooth. Hence, the phases are magnetically separated, increasing its reliability by diminishing the influence of one faulted phase over the remaining healthy one. The ratings for this machine are 48 Vdc, 500 rpm, 1.8 Nm, and 66.6 Hz of fundamental current frequency.

The sizing process is based on a generalized algorithm that one can use for any type of PMSM indifferent of the number of phases, slots, or pole pairs. Neglecting the leakage reactance, the output power can be a computed function of the estimated efficiency (η), the number of phases (n_{ph}), the peak value of the mmf (E_{max}), and the peak value of the phase current (I_{max}):

$$\begin{aligned} P_{out} &= \eta \cdot n_{ph} \cdot E_{\max} \cdot I_{\max} \\ Q_R &= n_{div} \cdot N_{mS} - n_{div} \end{aligned} \tag{21}$$

In Eq. (22), the peak value of the mmf can be a computed function of its coefficient (k_E), the number of turns per phase (N_t), the air-gap flux density (B_{gap}), the active stack length (L_m), the supply current frequency (f_s), and the number of pole pairs (p):

$$E_{\max} = k_E \cdot N_t \cdot B_{gap} \cdot L_m \cdot f_s / p \tag{22}$$

Defining the geometric coefficient, $k_L = L_m/D_{gap}$, the current coefficient $k_i = I_{max}/I_{rms}$, and defining the phase load in ampere-turns:

$$A_t = 2/\pi \cdot N_t \cdot I_{rms}/D_{gap} \tag{23}$$

One can compute the air-gap mean diameter, just like for the SRM in Eq. (2), using

$$D_{gap} = \sqrt[3]{\frac{2 \cdot p \cdot P_{out}}{\pi \cdot n_{ph} \cdot A_t \cdot k_E \cdot k_i \cdot k_L \cdot \eta \cdot B_{gap} \cdot f_s}} \tag{24}$$

Based on the value computed with Eq. (23), it is possible to size all the rest of geometrical dimensions of the machine, based on classical models. The resulting air-gap flux density can be determined with

$$B_{gap} = \frac{h_m \cdot B_{rm}}{\frac{D_{gap}}{2} \cdot \left(\ln\left(\frac{R_{si}-gap}{R_{cr}}\right) + \ln\left(\frac{R_{si}}{R_{si}-gap}\right)\right)} \tag{25}$$

where (h_m) is the length of the permanent magnet, (B_{rm}) is the permanent magnet's remanent flux density, (R_{si}) is the stator inner diameter, (R_{cr}) is the rotor core diameter, and (gap) is the air-gap length. The rest of the electromagnetic parameters can be computed based on the air-gap flux density. The phase emf, proportional with the frequency and the number of turns,

the flux linkage, and the demagnetization coefficient (k_d) given by the permanent magnet's specifications (around 0.8–0.9) can be determined:

$$E_{ph} = \sqrt{2} \cdot \pi \cdot f_s \cdot N_t \cdot k_{ws} \cdot \Psi_m \cdot k_d \tag{26}$$

where (k_{ws}) is the winding coefficient, specific for the used winding type.

The PMSM's speed is given proportional with the emf and inverse proportional with its coefficient:

$$\begin{aligned} \Omega &= E_{ph}/k_E \\ T_m &= P_{out}/\Omega \end{aligned} \tag{27}$$

The developed mean torque is a computed function of the output power and the machine speed. All that remain now, having the machine designed, are to validate its operation both in simulations and in laboratory measurements.

3.2. The operation of the nine-phase PMSM

Before proceeding to validation of the machine, it is important to describe the dedicated electronic power converter attached to the machine and the concept of its control, in order for the reader to be able to understand the choice of the system's assemblies.

One solution for a converter to be fault tolerant is to create H bridges around each phase and control them independently. However, such a structure is very expensive, and at the same time, it is quite difficult to create its proper control. A cost-effective solution is to build a converter like the one depicted in **Figure 11**. The concept of the inverter is to divide the nine phases of the machine in three groups of three phases [24, 25]. Each group of three phases forms a star connection obtaining a neutral point that is connected to an additional inverter leg. This is operated only if fault occurs on one of the three phases of one star. By this, the machine can be operated even if three phases are faulted, with only one of the three sets of phases. Moreover, even if one of the last three remaining phases is faulted, the machine can continue its operation with only two of the nine phases.

In case the fault occurs on one phase of a group of 3, this is automatically isolated by keeping the power switches around it open. However, due to the imbalance, the zero sequence current is not null anymore. To keep the balance, this current, (I_0), has to reach the supply. This is where the additional leg is operated, ensuring that the zero sequence current reaches the supply and by this, practically, the remaining healthy phases keep their balance driving normal currents to the machine.

In order to have a better understanding of the phenomenon, the discussion that follows will be held on a system of three-phase currents. However, as practically the three-phase machine has 3×3 phases, the concept is extendable applying the same procedures.

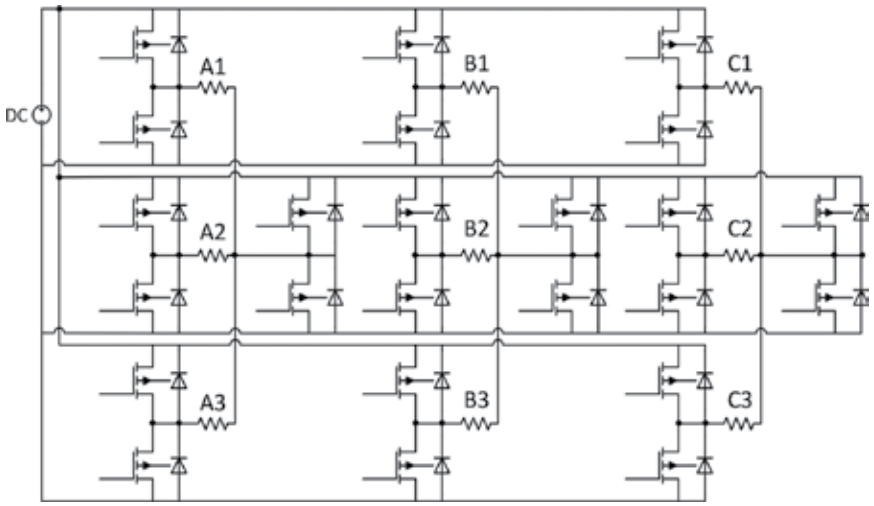


Figure 11. The nine-phase fault-tolerant inverter structure.

In a three-phase system, the currents in a rotating reference frame are governed by

$$\begin{aligned} I_a &= I_d \cdot \sin(\omega t) + I_q \cdot \cos(\omega t) + I_0 \\ I_b &= I_d \cdot \sin\left(\omega t - 2\frac{\pi}{3}\right) + I_q \cdot \cos\left(\omega t - 2\frac{\pi}{3}\right) + I_0 \\ I_c &= I_d \cdot \sin\left(\omega t + 2\frac{\pi}{3}\right) + I_q \cdot \cos\left(\omega t + 2\frac{\pi}{3}\right) + I_0 \end{aligned} \tag{28}$$

Writing those in fixed reference frame will give

$$\begin{aligned} I_a &= \frac{3}{2} I_\alpha + \frac{\sqrt{3}}{2} I_\beta \\ I_b &= \sqrt{3} I_\beta \\ I_c &= 0 \end{aligned} \tag{29}$$

The electromagnetic torque in general relation is expressed as

$$T_m = \frac{3}{2} \cdot p \cdot \left(I_q \cdot \Psi_m + \left(L_d - L_q\right) \cdot I_d \cdot I_q\right) \tag{30}$$

In relation to Eq. (30), it can be seen that to maintain a constant torque, the *dq* components of the currents must remain the same. The link between fixed and rotating referenced frame currents is given by

$$\begin{aligned} I_\alpha &= I_d \cdot \cos(\omega t) - I_q \cdot \sin(\omega t) \\ I_\beta &= I_d \cdot \sin(\omega t) + I_q \cdot \cos(\omega t) \end{aligned} \tag{31}$$

Hence, substituting Eq. (31) into Eq. (29) gives the three-phase current law needed to keep the developed electromagnetic torque constant:

http://dx.doi.org/10.5772/67354

$$\begin{aligned} I_a &= \sqrt{3}I_d \cdot \sin\left(\omega t + \frac{2\pi}{3}\right) + \sqrt{3}I_q \cdot \cos\left(\omega t + \frac{2\pi}{3}\right) \\ I_b &= \sqrt{3}I_d \cdot \sin(\omega t) + \sqrt{3}I_q \cdot \cos(\omega t) \\ I_c &= 0 \end{aligned} \tag{32}$$

Eq. (32) proves that the machine can keep the developed torque constant only if the currents are increased with $\sqrt{3}$. However, if the currents are not increased, in case of fault, the machine will develop only 2/3 of the rated torque. This concept demonstrates that, in case of faults, the machine can be supplied to develop the same torque by controlling the magnitude of the *dq* currents. It has to be mentioned that for such applications, the winding must be sized properly, to be able to handle increased currents without the risk of burnout.

Validation both in simulations and in laboratory experiments will be focused on injecting currents in the neutral point using the additional inverter leg, to compensate the fault occurred in each star connection of the machine.

Flux2D software, a software based on finite element analysis, was used to validate the machine via simulations.

In **Figure 12** the normal and one-phase fault conditions are depicted for the nine-phase PMSM. As it can be seen for the normal condition, all the nine currents are present developing torque, reaching 1.8 Nm, the rated value. In case of fault, this was set to occur at 0.01 s; it can be seen that the lack of one phase creates torque ripples and its RMS value diminishes to 1.6 Nm. In order to prove the concept, no currents will be increased, highlighting that the value of the torque is diminished with 2/3 for each phase.

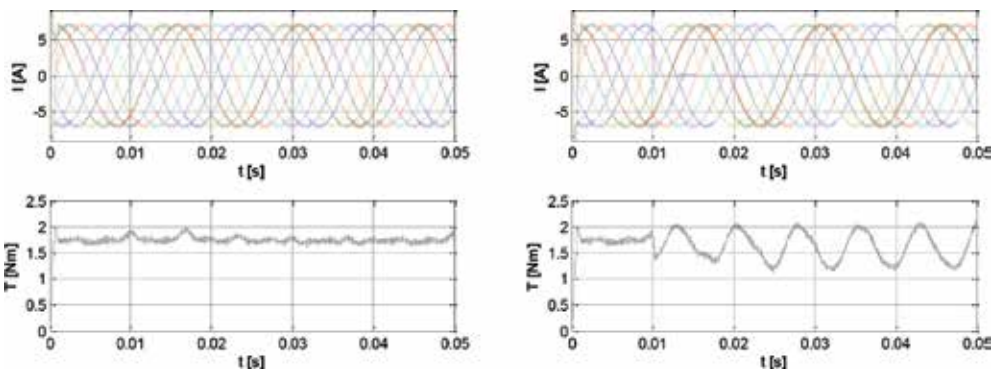


Figure 12. Simulated result in healthy (left) and one-phase fault (right).

Figure 13 depicts several fault condition tests that were performed. When two phases from two star connections are lacking (**Figure 13**, top-left), the torque ripples increase even more, and the RMS value reaches 1.5 Nm. If a third phase (**Figure 13**, top-right), from the third star connection, is faulted, the ripples increase again, and the RMS value of the torque decreases to 1.35 Nm. The worst case scenario is depicted in **Figure 13** (bottom). Here, two phases from one star and one from each of the other stars are faulted. Hence, only five phases remain

operational in the machine. High torque ripples are reached in this case, and the torque developed barely passes 1 Nm as mean value.

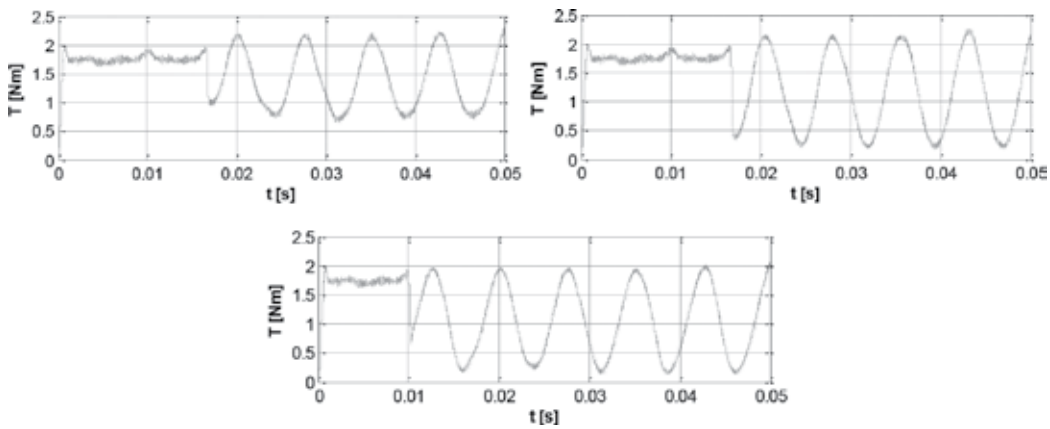


Figure 13. Simulated results in faulted conditions: two phases (top-left), three phases (top-right), and four phases (bottom).

The experimental validation of the nine-phase PMSM was done with the test bench depicted in **Figure 14**. Its main assemblies are the PMSM, a torque meter, an induction machine as load, the complex electronic converter, nine current sensors, a power supply, and a dSPACE 1103 card for the control [16].

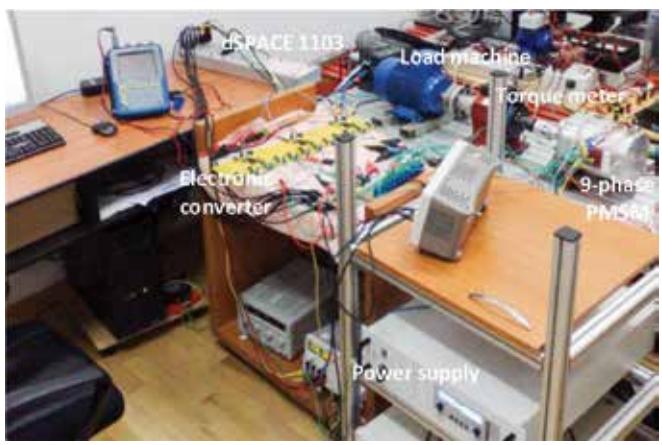


Figure 14. The nine-phase PMSM test bench.

In order to highlight the benefit of using the additional leg of each star connection, a debate will be focused on only one of the three star connections, the phenomenon being the same for the entire machine. The three-phase currents and the developed torque in healthy condition are depicted in **Figure 15** (top). The amplitude of the currents and torque reaches rated values; these are being measured at the rated speed. The deformation of the currents is given by low

http://dx.doi.org/10.5772/67354

switching frequency. The developed torque is about 1.72 Nm as RMS value, quite close to the one obtained in simulations. In case of a phase fault of one star (obtained by forcing open circuit), without engaging the fourth leg, the currents become shifted with 180°, like in **Figure 15** (middle). However, engaging the fourth leg, this will complete the switching table of a three-phase system, keeping the currents shifted with the proper angle.

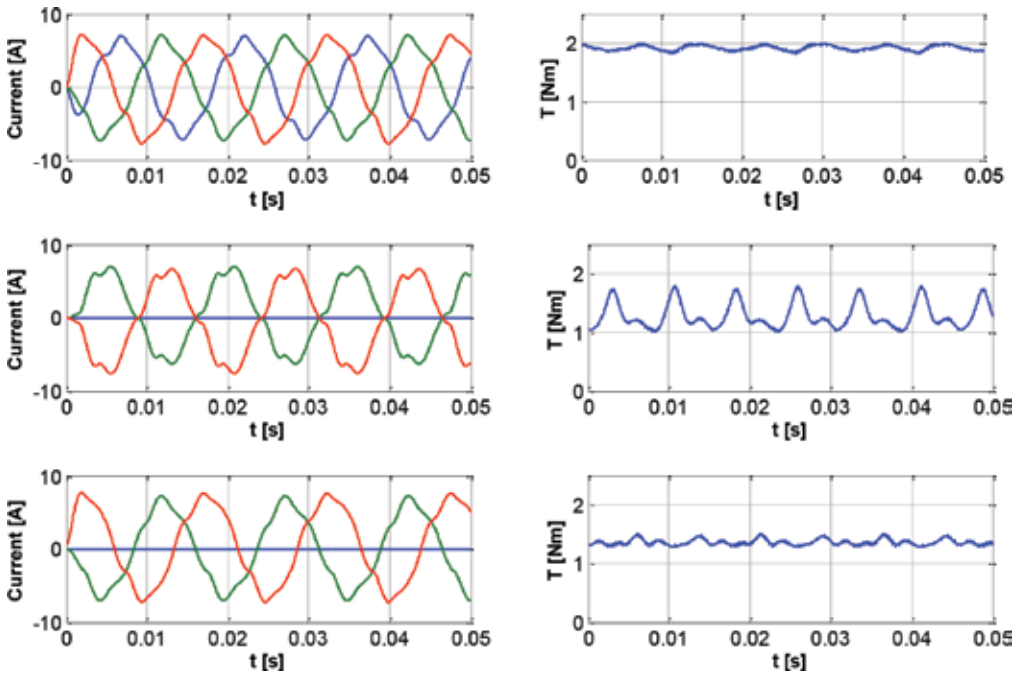


Figure 15. The currents (left) and the instantaneous torque (right) for healthy machine (top) and in one-phase fault without (middle) and with (bottom) the contribution of the fourth leg.

The torque is diminished in this case, but the ripples are much lower than in the case when the fourth leg was not engaged. The RMS torque obtained was 1.62 Nm, quite close to the one resulted simulations wise.

This analysis proves the benefits of using such a configuration of an additional fourth leg correspondence for each group of three phases in star connection. This operational concept is extendable for any number of phases. As the machine in study has nine phases, the same fault conditions that were tested in simulations will be applied in the experimental validation too.

In **Figure 16**, nine-phase currents are plotted for healthy machine operation. The reason of the distortion was already explained in the above paragraphs. To compare the results with those obtained in simulation, firstly two phases are opened from two different star connections. It can be seen in **Figure 16** (top-right) that for this condition the torque ripples are higher than in the case of only one phase fault. The RMS torque developed was about 1.42 Nm. Another fault condition was to open three phases, each from a different star connection. The result plotted in **Figure 16** (bottom-left) shows not only the high torque ripples but also some instability of the

characteristic. In this case about 1.28 Nm RMS torque was developed. The worst case scenario is having two opened phases from the first star connection and other two phases from each of the remaining stars. The torque depicted in **Figure 16** (bottom-right) presents the highest torque ripples and an RMS torque of 1.18 Nm. Comparing the results from simulations and from experimental measurements in the same operation conditions, there are quite consistent differences between them. The reasons are due to the fact that the analysis conditions were quite different. The switching frequency during experiments was limited to the computation power of dSPACE. By this, the shape of the currents included a certain content of harmonics that has direct impact on the shape of the torque. The rotating assemblies of the test bench had consistent inertia adding together one of the loads of the mechanical coupling and of the PMSM's itself. Overall, the analysis highlighted the benefits of using a fractional-slot design for the machine to diminish the torque ripples supplied by a special electronic converter able to isolate and compensate the faults. The level of fault tolerance was proven by the fact that the machine is able to operate with only five of its nine phases, continuing to develop quite a reasonable quantity of torque.

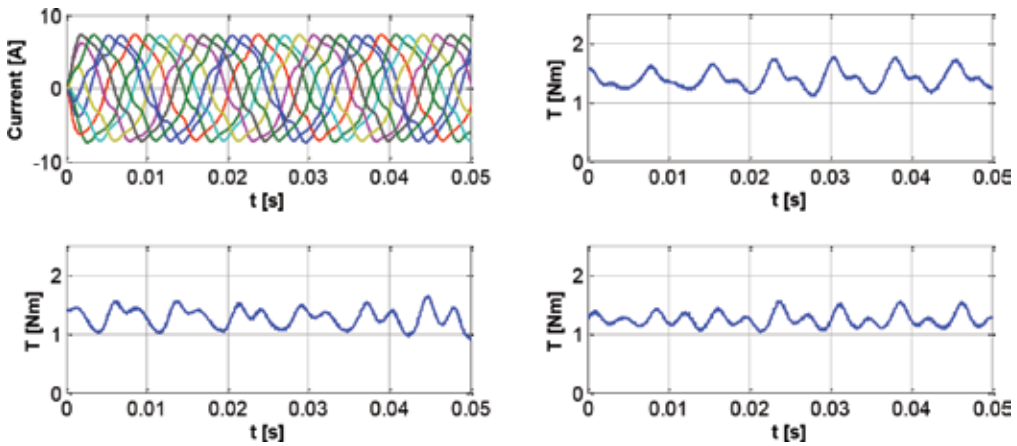


Figure 16. The nine-phase currents (top-left) and the instantaneous torque for fault on two star connections (top-right), fault on three star connections (bottom-left), and fault on two phases of a star and fault of one phase of each of the other two stars (bottom-right).

4. Conclusions

In the present chapter, fault-tolerant electrical machines and drives are presented for particular applications in the light electric vehicle industry. The goal was to analyze the most used machine nowadays in this area such as the switched reluctance and the permanent-magnet synchronous machines. The latter one was designed in a multiphase structure, having nine-phases, each three of them connected in a star connection. Operating this machine with a dedicated special architecture electronic converter, one can reach continuous operation despite open-phase conditions. Numerical simulations and experimental tests prove the operational skills of the machine.

The fault-tolerant SRM was designed in a novel concept, to have a modular stator, each module caring independent coils. Each diametrically opposed two coils form one stator phase, but each coil is supplied independently from a half-H bridge. For this machine also numerical and experimental studies proved the integrity and the usefulness of the machine.

The presented structures are part of a step forward in the field of high reliability electrical machines. These, controlled with the proposed electronic converters and adding intelligent control strategies, are able to reach high fault-tolerant capabilities. On the other hand, considering the price of implementing such structures, it is clear that these are higher than the classical ones. However, fault-tolerant systems are by default high-cost designs, and these costs are increasing with their complexity and reliability.

Author details

Mircea Ruba

Address all correspondence to: mircea.ruba@emd.utcluj.ro

Technical University of Cluj Napoca, Romania

References

[1] B. Tabbache, A. Kheloui, M. Benbouzid, A. Mamoune, D. Diallo. "Research on Fault Analysis and Fault-Tolerant Control of EV/HEV Powertrain," in IEEE ICGE 2014, Mar 2014, Sfax, Tunisia. pp. 284–289.

[2] Y. Zhang, J. Jiang. "Bibliographical review on reconfigurable fault-tolerant control systems," ELSEVIER, Annual Reviews in Control, vol. 32, 2008, pp. 229–252. DOI: 10.1016/j.arcontrol.2008.03.008

[3] Y. Hu, C. Gan, et al. "Flexible fault-tolerant topology for switched reluctance motor drives," IEE Transactions on Power Electronics, vol. 31, no. 6, 2016, pp. 4654–4668.

[4] N. S. Gameiro, A. J. Marques Cardoso. "A new method for power converter fault diagnosis in SRM drives," IEEE Transactions on Industrial Applications, vol. 48, no. 2, 2012, pp. 653–662.

[5] B. Lequesne, S. Gopalakrishnan, A. M. Omekanda. "Winding short circuits in the switched reluctance drive," IEEE Transactions on Industrial Applications, vol. 41, no. 5, 2005, pp. 1178–1184.

[6] W. Ding, Y. Liu, Y. Hu. "Performance evaluation of a fault-tolerant decoupled dual-channel switched reluctance motor drive under open-circuits," IET Electric Power Applications, vol. 8, no. 4, 2014, pp. 117–130.

[7] C. J. Gajanayake, B. Bhangu, et al. "Fault tolerant control method to improve the torque and speed response in PMSM drive with winding faults," 2011 IEEE Ninth International Conference on Power Electronics and Drive Systems (PEDS), 5–8 Dec. 2011. DOI: 10.1109/PEDS.2011.6147371

[8] O. Wallmark, L. Harnefors, O. Carlson. "Control algorithms for a fault-tolerant PMSM drive," IEEE Transactions on Industrial Electronics, vol. 54, no. 4, 2007, pp. 1973–1980. DOI: 10.1109/TIE.2007.895076

[9] N. K. Nguyen, F. Meinguet, E. Semail, X. Kestelyn. "Fault-tolerant operation of an open-end winding five-phase PMSM drive with short-circuit inverter fault," IEEE Transactions on Industrial Electronics, vol. 63, no. 1, 2016, pp. 595–605. DOI: 10.1109/TIE.2014.2386299

[10] H. Berriri, M. Naouar, I. Slama-Belkhodja. "Easy and Fast Sensor Fault Detection and Isolation Algorithm for Electrical Drives," IEEE Transactions on Power Electronics, vol. 27, no. 2, Feb. 2012, pp. 490–499. ISSN: 1941-0107

[11] I. A. Viorel, S. Larisa, I. F. Soran, "Analytical flux linkage model of switched reluctance motor," Romanian Journal of Technical Sciences – Electrotechnical And Energy Series, vol. 54, no. 2, 2009, pp. 139–146.

[12] G. Henneberger, I. A. Viorel. "Variable Reluctance Electrical Machines." Shaker Verlag, Aachen (Germany), 2001.

[13] A. Radun. "Design considerations for the switched reluctance motor," Proceedings of IEEE Transactions on Industrial Applications, vol. 31, no. 5, 1995, pp. 1079–1087.

[14] R. Krishnan. "Switched Reluctance Motor Drives – Modeling, Simulation, Analysis, Design, and Applications." Industrial Electronics Series, CRC Press, Alabama, USA, 2001. ISBN 0-8493-0838-0.

[15] M. Ruba, I. A. Viorel, L. Szabó. "Modular stator switched reluctance motor for fault tolerant drive systems," IET Electric Power Applications, vol. 7, no. 3, 2013, pp. 159–169. ISSN 1751-8660.

[16] G. Oriti, A. L. Julian, T. A. Lipo. "An inverter/motor drive with common mode voltage elimination," in IEEE Industry Application Society Annual Meeting, New Orleans, Louisiana, 5–9 October 1997, pp. 587–592.

[17] M. Jones, S. N. Vukosavic, E. Levi, A. Iqbal. "A six-phase series connected two-motor drive with decoupled dynamic control," IEEE Transactions on Industry Applications, vol. 41, no. 4, 2005, pp. 1056–1066.

[18] M. B. R. de Correea, C. B. Jacobina, E. R. C. da Silva, E. M. N. Lima. "An induction motor drive system with improved fault tolerance," IEEE Transactions on Industry Applications, vol. 37, no. 3, 2001, pp. 873–879.

[19] N. Bianchi, S. Bolognani, M. Zigliotto, M. Zordan. "Innovative remedial strategies for inverter faults in IPM synchronous motor drives," IEEE Transactions on Energy Conversion, vol. 18, no. 2, 2003, pp. 306–314.

[20] W. Ouyang, T. A. Lipo. "Modular permanent magnet with fault tolerant capability," in 24th Annual IEEE Applied Power Electronics Conference and Exposition, 2009, pp. 930–937.

[21] C. H. Sneessens, T. Labbe, F. Baudart, F. Labrique, E. Matagne. "Modelling and torque control of a five-phase permanent magnet synchronous motor using tooth concentrated winding technology," in Electromotion 2009, 8th International Symposium on Advanced Electromechanical Motion Systems & Electric Drives Joint Symposium, 2009, pp. 1–6.

[22] M. J. Duran, F. Barrero, S. Toral. "Multi-phase space vector pulse width modulation: applications and strategies," in International Conference on Renewable Energies and Power Quality, 2007, paper ID 341.

[23] A. Akrad, M. Hilairet, D. Diallo. "Design of fault-tolerant controller based on observers for a PMSM drive," IEEE Transactions on Industrial Electronics, vol. 58, no. 4, 2011, pp. 1416–1427.

[24] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki. "Diagnosis and Fault-Tolerant Control." Springer, 2nd ed., 2006, XIX, 672 p. 270 illus.

[25] D. Fodorean, F. Jurca, M. Ruba, D. C. Popa. "Motorization Variants for Light Electric Vehicles—Design, Magnetic, Mechanical and Thermal Aspects." Alma Mater, Cluj Napoca, Romania, 2013. ISBN 978-606-504-160-8

Chapter 9

Fault Detection and Isolation of Nonlinear Systems with Generalized Hamiltonian Representation

Luis Humberto Rodriguez-Alfaro,
Efrain Alcorta-Garcia, Cornelio Posadas-Castillo and
David Alejandro Diaz-Romero

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/68084

Abstract

The problem of fault diagnosis in a class of nonlinear system is considered. Systems that can be written in the so-called Generalized Hamiltonian Representation (which is equivalent to an Euler-Lagrange representation) are studied, and a model-based observer approach for this class of systems is developed. The main advantage of the proposed approach is the facility to design the required observers, which take advantage of the system structure given by the Hamitonian representation. In order to show the proposed schema, a model of a permanent magnet synchronous machine is revised and the fault diagnosis schema presented. Simulation results confirm the effectivity of the proposed schema.

Keywords: fault diagnosis, Hamiltonian systems, nonlinear systems, observers, fault isolation

1. Introduction

Safety operation and reliability of industrial processes are highly prized by the contemporary society. A key to achieve safety and reliability in industrial processes is through the use of diagnosis and fault-tolerant control algorithms. Note that a fault is understood as a change of a parameter out of the tolerance limits. Physical systems are liable to potentially harmful fault events, which could cause a negative effect on the system functionality, as well as underperformance. Faults can be originated by diverse reasons, for example, natural wear caused by common use, aging, use under stress conditions and so on. The importance of detecting and isolating the fault occurrence in a system lies in the possibility to reduce the maintenance and/or dead-time for repairing on a production line.

There exist a lot of results related to fault diagnosis for linear systems, as it can be seen in the literature, for example, in Refs. [1–6], among others. For the case of nonlinear systems, there are also some available solutions based on diverse model structures, see, for example, Refs. [7, 8]. Fault diagnosis in nonlinear systems has been considered in Ref. [9], where the solution is based on a geometric approach, and the conditions of existence are not easily satisfied. Other approaches consider Lipschitz-type nonlinear systems together with an observer-based method [10, 11]. Ref. [12] is related to the problem of fault estimation for a class of switched nonlinear systems of neutral type, where the problem formulated as an H_∞ filtering is solved using a switched observer-based fault estimator. In Ref. [13], the fault diagnosis is made for a class of bilinear systems considering only the case of faults on the actuator.

In Ref. [14], an unknown input observer (UIO) for a class of nonlinear state-affine systems for fault diagnosis is proposed. By using sum-of-squares (SOS) theory and Lie geometry as the main tools, the rank constraint in the traditional UIO approach is relaxed and the design procedure simplified, especially for the case of nonlinear polynomial systems. In Ref. [15, 16], an approach to fault detection and isolation for the class of nonlinear systems with linear parameter varying (LPV) systems is shown. A different idea is to use a energy index in the diagnosis process, as in Ref. [17]. In Ref. [18], an algorithm for the diagnostics of nonlinear systems is presented where the solution is based on the estimation of the system parameters using the nonlinear response. The use of a bank of high-order sliding mode observers has been proposed in Ref. [19].

From the above discussion, it is clear that even if some approaches are available to settle the fault diagnosis problem, in general there is no systematic way to design it (a model-based or an observer-based approach), because of the difficulty to design an observer for nonlinear systems even if the system is known. The available solutions consider a specific class of nonlinear systems, but each of these class of systems is more related to some mathematical (or system) properties and not necessarily to a wide class of systems from a practical point of view. Systems in Hamiltonian representation form represent a wide range of physical systems considering the relationship between Euler-Lagrange and Hamiltonian systems [20–22].

In this chapter, a solution to the problem of fault detection and isolation applying the observer-based residual generation method is proposed. The class of nonlinear systems considered includes all systems, which admit a generalized Hamiltonian representation. The proposed solution begins with a mathematical nonlinear model of a system with faults. A nonlinear decoupling is applied to the faulty system in order to obtain a set of subsystems with sensibility to a particular fault or group of faults. Then, each subsystem is represented in a generalized Hamiltonian form, for which, a nonlinear observer is designed. Using the nonlinear observer, the residual generator is designed for each subsystem. One contribution of this work is the systematic way for residual generator design (an observer-based approach with weak design requirements). Note that the observer-based approach is guaranteed because of the Hamiltonian representation. Fault detection and isolation follow from the residual analysis. The structure of the Hamiltonian system representation is exploited to guarantee the residual existence for each subsystem. The approach is then applied to the model of a permanent magnet synchronous machine with additive faults. The faults are detected and isolated conveniently, showing the effectiveness of the proposed approach.

http://dx.doi.org/10.5772/68084

2. Generalized Hamiltonian representation of a system

Consider a nonlinear system described in general form as follows:

$$\begin{aligned} \underline{\mathbf{x}} &= \mathbf{f}(\mathbf{x},\mathbf{u}), \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}), \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{u} \in \mathbb{R}^r$ is the input vector, $\mathbf{y} \in \mathbb{R}^m$ represents the output vector, and the function $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^r \rightarrow \mathbb{R}^n$ associates with each value of $\mathbf{x}$ and $\mathbf{u}$ a corresponding n dimensional vector.

A special class of Generalized Hamiltonian representation is defined by Sira-Ramierez and Cruz-Hernandez [23],

$$\begin{aligned} \underline{\mathbf{x}} &= [\mathbf{J}(\mathbf{x}) + \mathbf{S}] \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \mathbf{G}\mathbf{u}, \\ \mathbf{y} &= \mathbf{C} \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}}, \end{aligned} \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes the state vector, $\mathbf{u} \in \mathbb{R}^r$ is the input vector, $\mathbf{G} \in \mathbb{R}^{n \times r}$ is a constant matrix, $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^n$ denotes a vector that contains the nonlinearities, $\mathbf{y} \in \mathbb{R}^m$ denotes the output vector, and $\mathbf{C} \in \mathbb{R}^{m \times n}$ is a constant output matrix. Some nonlinear systems such as these described by Eq. (1) can be represented by Eq. (2) if satisfies the following conditions: There exists a smooth energy function $H(\mathbf{x})$ that is positive definite in $\mathbb{R}^n$ and described by:

$$H(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{M} \mathbf{x}, \tag{3}$$

the column gradient vector denoted by $\frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{M}\mathbf{x}$ can be obtained using Eq. (3), where $\mathbf{M} \in \mathbb{R}^{n \times n}$ must be a symmetric matrix constant and positive definitive, $\mathbf{J}(\mathbf{x}) \in \mathbb{R}^{n \times n}$ must be satisfied for all $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{S} \in \mathbb{R}^{n \times n}$ is a constant symmetric matrix,

$$\mathbf{J}(\mathbf{x}) = -\mathbf{J}^T(\mathbf{x}), \mathbf{S} = \mathbf{S}^T. \tag{4}$$

These conditions allow that a wide set of nonlinear systems can be brought to a generalized Hamiltonian representation, such as electromechanical systems, electric systems, mechanical systems, etc.

In the generalized Hamiltonian representation, the additive faults can be represented as in Eq. (5), where these appear as additional inputs (unknown inputs).

$$\begin{aligned} \underline{\mathbf{x}} &= \mathbf{J}(\mathbf{x}) \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{S} \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \mathbf{G}\mathbf{u} + \mathbf{N}(\Delta f), \\ \mathbf{y} &= \mathbf{C} \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{Q}(\Delta f), \end{aligned} \tag{5}$$

where $\mathbf{N}(\Delta f) \in \mathbb{R}^n$ and $\mathbf{Q}(\Delta f) \in \mathbb{R}^m$ represent the additive faults of the system.

3. Fault detection and isolation

In a general sense, if the fault diagnosis consists in the detection of a fault, then it is called fault detection(FD), and similarly, if the fault diagnosis consists in the detection and isolation of a fault, then it is called fault detection and isolation (FDI). The fault detection consists in determine the occurrence of faults in the functional units of the process, which leads to undesired behavior of the system, and the fault isolation consists in to classify the detected faults. The observer based fault diagnosis technique is a scheme of the model-based fault diagnosis approach. In this technique, the idea is to replace the process model by an observer which estimates the fault-free process outputs. The difference between the measured process variables and the estimated process variables defines the residual. The fault effect is contained in the measured process variables. Thus, a residual signal includes the fault effect. Ideally, if the residual is different from zero then a fault has occurred, otherwise the process is fault free. The residual generation allows to know the occurrence of faults, and the residual evaluation is necessary to extract the fault information. **Figure 1** shows a common diagnosis scheme.

In this contribution, a fault detection and isolation approach to nonlinear systems that admit a generalized Hamiltonian representation is considered. The proposed approach follows the classical procedure of fault diagnosis: First, a fault decoupling in order to get subsystems with sensibility to a specific fault is developed. Second, an observer-based residual generator for each subsystem is designed. Third, a residual analysis is performed to determine which functional unit has failed.

Figure 2 shows the proposed fault detection and isolation scheme, where C_n is the nominal control, Σ_H is a system in Hamiltonian representation, and the diagnostic block contains the observer and the residual generator.

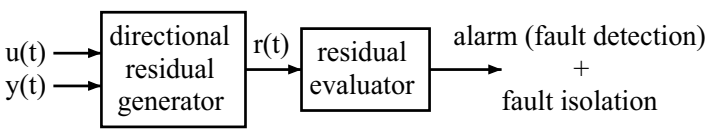


Figure 1. Fault diagnosis scheme.

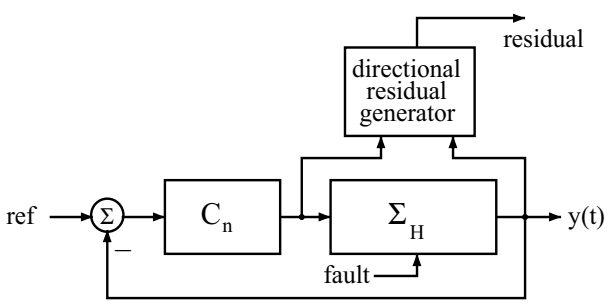


Figure 2. Diagnostic scheme.

http://dx.doi.org/10.5772/68084

The fault decoupling consists in to define a transformation over the system in order to get a subsystem with sensibility to a fault or a set of faults, and this subsystem is coupled with a fault and decoupled from the rest of possible faults. There are some works on the analysis and synthesis of these transformations, see, for example, Refs. [9, 24, 25].

For the case of the generalized Hamiltonian representation with faults, Eq. (5) considers the following nonlinear transformation

$$\zeta = \mathbf{T}(\mathbf{x}). \tag{6}$$

It is required that

$$\dot{\zeta} = \frac{\partial \mathbf{T}(\mathbf{x})}{\partial \mathbf{x}} \dot{\mathbf{x}}, \tag{7}$$

$$\dot{\zeta} = \frac{\partial \mathbf{T}(\mathbf{x})}{\partial \mathbf{x}} \left[\mathbf{J}(\mathbf{x}) \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{S} \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \mathbf{G}\mathbf{u} + \mathbf{N}(\Delta f) \right], \tag{8}$$

and the transformation $\mathbf{T}(\mathbf{x})$ be selected in such a way that the resulting transformed system has the desired fault sensibility, that is, suppose $\mathbf{N}(\Delta f) = [n_1(\Delta f)\ n_2(\Delta f) \cdots n_l(\Delta f)]$ where $\overline{\mathbf{N}}(\Delta f)$ represents the columns associated with the faults that requires no to affect a specific subsystem and $\overline{\overline{\mathbf{N}}}(\Delta f)$ are the columns related to the faults that are required to affect the subsystem. With $\frac{\partial \mathbf{T}(\mathbf{x})}{\partial \mathbf{x}} \overline{\mathbf{N}}(\Delta f) = 0$ and $\frac{\partial \mathbf{T}(\mathbf{x})}{\partial \mathbf{x}} \overline{\overline{\mathbf{N}}}(\Delta f) \neq 0$. In Ref. [24], it can be found details about the existence of this transformation.

Assumption 1. *Consider the system Eq. (1) in generalized Hamiltonian representation with faults as in Eq. (2) as well as the nonlinear transformation $T(x)$ satisfying decoupling requirements. Also, the transformed system (decoupled) can be represented in the Hamiltonian form given by Eq. (2).*

For some examples, at least, the assumption is satisfied and consequently, a systematic way to fault isolation is obtained. At the moment, we do not have a result on the characterization of the class of systems for which the assumption is satisfied. The resulting decoupled system is represented in a Hamiltonian form.

$$\begin{aligned} \dot{\zeta} &= \left[\overline{\mathbf{J}}(\zeta) + \overline{\mathbf{S}} \right] \frac{\partial \overline{H}(\zeta)}{\partial \zeta} + \overline{\mathbf{F}}(\zeta) + \frac{\overline{\mathbf{G}}}{\overline{\mathbf{u}}} + \overline{\overline{N}}(\Delta f), \\ \mathbf{y}_\zeta &= \overline{\mathbf{C}} \frac{\partial \overline{H}(\zeta)}{\partial \zeta}, \end{aligned} \tag{9}$$

where $\zeta \in \mathbb{R}^{n_\zeta}$ denotes the state vector, $\overline{\mathbf{J}}(\zeta) \in \mathbb{R}^{n_\zeta \times n_\zeta}$, $\overline{\mathbf{S}} \in \mathbb{R}^{n_\zeta \times n_\zeta}$, $\mathbf{u} \in \mathbb{R}^{r_\zeta}$ is the input vector, $\overline{\mathbf{G}} \in \mathbb{R}^{n_\zeta \times r_\zeta}$ is a constant matrix, $\overline{\mathbf{F}}(\zeta) \in \mathbb{R}^{n_\zeta}$ denotes a vector that contains the nonlinearities, $\mathbf{y}_\zeta \in \mathbb{R}^{m_\zeta}$ denotes the output vector, and $\overline{\mathbf{C}} \in \mathbb{R}^{m_\zeta \times n_\zeta}$ is a constant output matrix.

After a subsystem has been determined, the next step is to design an observer for each subsystem. From the decoupled subsystem in the generalized Hamiltonian representation Eq. (9), an observer can be designed as follows [26]:

$$\begin{aligned}\dot{\hat{\zeta}} &= \overline{\mathbf{J}}(\hat{\zeta})\frac{\partial\overline{H}(\hat{\zeta})}{\partial\hat{\zeta}} + \overline{\mathbf{S}}\frac{\partial H(\hat{\zeta})}{\partial\hat{\zeta}} + \overline{\mathbf{F}}(\hat{\zeta}) + \overline{\mathbf{G}}\overline{\mathbf{u}} + \mathbf{K}(\mathbf{y}_\zeta - \hat{\mathbf{y}}_\zeta),\\ \hat{\mathbf{y}}_\zeta &= \overline{\mathbf{C}}\frac{\partial\overline{H}(\hat{\zeta})}{\partial\hat{\zeta}},\end{aligned} \tag{10}$$

where $\mathbf{K}\in\mathbb{R}^{n\times m}$ is the observer gain, $\zeta\in\hat{\mathbb{R}}^n$ is the estimated state, $\hat{\mathbf{y}}_\zeta\in\mathbb{R}^m$ is the estimated output calculated in terms of $\hat{\zeta}$, $\frac{\partial H(\hat{\zeta})}{\partial\hat{\zeta}} = \mathbf{M}\hat{\zeta}$ is the gradient vector with $\mathbf{M}\in\mathbb{R}^{n\times n}$ as a symmetric positive definite matrix.

For this observer, the conditions design is described in the following Theorem:

Theorem 1. *The state* $\mathbf{x}$ *of the nonlinear system in the generalized Hamiltonian representation Eq. (9) can be globally, exponentially, asymptotically estimated by the observer Eq. (10), if the pair (C,S) is observable or at least detectable and the matrix*

$$\mathbf{M}\left[\mathbf{S} - \frac{1}{2}(\mathbf{K}\mathbf{C} + \mathbf{C}^T\mathbf{K}^T)\right]\mathbf{M} + \Pi, \tag{11}$$

is negative definite. With $\Pi = \frac{1}{2}\left[\mathbf{M}\frac{\partial F(\mathbf{r})}{\partial\mathbf{x}} + \left(\frac{\partial F(\mathbf{r})}{\partial\mathbf{x}}\right)^T\mathbf{M}\right]$ *and* $\boldsymbol{\rho}$ *is a vector such that* $\boldsymbol{\rho}\in(\mathbf{x},\hat{\zeta})$.

◇◇◇

The proof of Theorem 1 is fully defined and explained in Ref. [26]. Then, for the decoupled system, a residual generator is defined as follows

Theorem 2. *For the decoupled nominal system (Eq. (9) with* $\overline{\overline{N}} = 0$*). The system*

$$\dot{\hat{\zeta}} = \overline{\mathbf{J}}(\hat{\zeta})\frac{\partial\overline{H}(\hat{\zeta})}{\partial\hat{\zeta}} + \overline{\mathbf{S}}\frac{\partial\overline{H}(\hat{\zeta})}{\partial\hat{\zeta}} + \overline{\mathbf{F}}(\hat{\zeta}) + \overline{\mathbf{G}}\mathbf{u} + \mathbf{K}\left(\mathbf{y}_\zeta - \overline{\mathbf{C}}\frac{\partial\overline{H}(\hat{\zeta})}{\partial\hat{\zeta}}\right), \tag{12}$$

$$\mathbf{r} = \mathbf{y}_\zeta - \overline{\mathbf{C}}\frac{\partial\overline{H}(\hat{\zeta})}{\partial\hat{\zeta}}, \tag{13}$$

is a directional residual generator if the pair $(\overline{C}, \overline{S})$ *is observable or at least detectable and the matrix*

$$\mathbf{M}\left[\overline{\mathbf{S}} - \frac{1}{2}(\mathbf{K}\overline{\mathbf{C}} + \overline{\mathbf{C}}^T\mathbf{K}^T)\right]\mathbf{M} + \Pi, \tag{14}$$

is negative definite. With $\Pi = \frac{1}{2}\left[\mathbf{M}\frac{\partial\overline{\mathbf{F}}(\mathbf{r})}{\partial\zeta} + \left(\frac{\partial\overline{\mathbf{F}}(\mathbf{r})}{\partial\zeta}\right)^T\mathbf{M}\right]$ *and* $\boldsymbol{\rho}$ *is a vector such that* $\boldsymbol{\rho}\in(\zeta,\hat{\zeta})$.

◇◇◇

http://dx.doi.org/10.5772/68084

Proof:

The proof of Theorem 2 is a consequence of the proof of Theorem 1.

◇◇◇

4. Application example

In this section, the results to apply in the **p**ermanent **m**agnet **s**ynchronous **m**achine (PMSM) the proposed approach for fault detection and isolation are presented. The closed loop system is used in the fault diagnosis analysis where any specific control law is used.

The PMSM mathematical model in the stationary reference frame $dq0$ (direct-quadrature-zero axes) is taken from Ref. [27] and is described by:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{G_m}\mathbf{u_m}, \tag{15}$$

where $\mathbf{x} = [i_d \; i_q \; \omega]^T$, $\mathbf{G_m} = diag\left[\frac{1}{L} \quad \frac{1}{L} \quad -\frac{1}{J_m}\right]$, $\mathbf{u_m} = diag[\, u_d \quad u_q \quad \tau_L \,]$ and

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} -\frac{R}{L} i_d + P\omega i_q 0.3cm \\ -\frac{R}{L} i_q - P\omega i_d - \frac{P\Phi}{L}\omega 0.3cm \\ \frac{3P\Phi}{2J_m} i_q - \frac{B}{J_m}\omega \end{bmatrix},$$

where B is the viscous friction coefficient, R is the stator resistance, L is the inductance, Φ is the flux linkage, P is the pole pairs, i_d and i_q are the electric currents on the direct and quadrature axis, respectively, u_d and u_q are the voltages on the direct and quadrature axes, respectively, ω is the rotor speed, J_m is the rotor inertia, and τ_L is the load torque.

In the fault diagnosis analysis, it is considered that the system is operating in nominal conditions, which implies that the system is in closed loop with any controller. In this case, a backstepping nonlinear control [22] is used in the PMSM.

In order to obtain the Hamiltonian representation Eq. (2) of the PMSM described by Eq. (15), a Hamiltonian energy function is defined as follows:

$$H(\mathbf{x}) = \frac{1}{2}\left(i_d^2 + Li_q^2 + \frac{2}{3}J_m\omega^2\right), \tag{16}$$

with a gradient vector

$$\frac{\partial}{\partial \mathbf{x}} H(\mathbf{x}) = \left[\, i_d \quad Li_q \quad \tfrac{2}{3}J_m\omega \,\right]^T, \tag{17}$$

$$\frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{M}\mathbf{x} \;\Rightarrow\; \mathbf{M} = diag\left[1 \quad L \quad \frac{2J_m}{3}\right], \tag{18}$$

where $\mathbf{M}$ is a symmetric, positive definite and constant matrix so that the Hamiltonian representation of the PMSM is as follows:

$$\dot{\mathbf{x}} = \mathbf{J}(\mathbf{x})\frac{\partial H}{\partial \mathbf{x}} + \mathbf{S}\frac{\partial H}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \mathbf{G}\mathbf{u}, \tag{19}$$

$$\mathbf{y} = \mathbf{C}\frac{\partial H}{\partial \mathbf{x}}, \tag{20}$$

where $\mathbf{x} = [i_d \; i_q \; \omega]^T$, $\mathbf{u} = [u_d \; u_q \; \tau_L]^T$,

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\dfrac{3P\Phi}{2J_m L} \\ 0 & \dfrac{3P\Phi}{2J_m L} & 0 \end{bmatrix},$$

$$\mathbf{S} = \begin{bmatrix} -\dfrac{R}{L} & 0 & 0 \\ 0 & -\dfrac{R}{L^2} & 0 \\ 0 & 0 & -\dfrac{3B}{2J_m^2} \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/L & 0 \\ 0 & 0 & \dfrac{3}{2J_m} \end{bmatrix},$$

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} P\omega i_q \\ -P\omega i_d \\ 0 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \dfrac{1}{L} & 0 & 0 \\ 0 & \dfrac{1}{L} & 0 \\ 0 & 0 & -\dfrac{1}{J_m} \end{bmatrix}.$$

Solving the Hamiltonian representation Eq. (19) for each of the state equations, the same model described by Eq. (15) is obtained, so that the Hamiltonian representation is correct since it fulfills the conditions Eq. (4).

An intermittent connection, signal lost or signal offset are some of the sensor faults prone to occur in electrical machines [28], the control objective is affected mainly by first and second faults. The nominal value of the load torque is known, an unknown change in this parameter is considered as an additive fault. The PMSM may occur faults on elements such as sensors, actuators and components. The following additive faults are considered in this contribution: $\Delta\omega$ is a fault in the speed sensor, f_a is a fault in the control input, and $\Delta\tau_L$ is an unknown change in the load torque.

When these faults are considered, the Hamiltonian representation of the PMSM is as follows:

$$\begin{aligned} \underline{\mathbf{x}} &= \mathbf{J}(\mathbf{x})\frac{\partial H}{\partial \mathbf{x}} + \mathbf{S}\frac{\partial H}{\partial \mathbf{x}} + \mathbf{F}(\mathbf{x}) + \mathbf{G}\mathbf{u} + \mathbf{N}(\mathrm{D}\mathbf{f}), \\ \mathbf{y} &= \mathbf{C}\frac{\partial H}{\partial \mathbf{x}} + \mathbf{Q}(\mathrm{D}\mathbf{f}), \end{aligned} \tag{21}$$

http://dx.doi.org/10.5772/68084

where **x**, **S**, **J**(**x**), **F**(**x**), **G** and **u** are the same as in the nominal case when there are no faults and

$$\mathbf{N}(\mathrm{D}\mathbf{f}) = \begin{bmatrix} f_a \\ f_a \\ \Delta\tau_L \end{bmatrix}, \mathbf{Q}(\Delta\mathbf{f}) = \begin{bmatrix} 0 \\ 0 \\ \Delta\omega \end{bmatrix}.$$

Once defined the mathematical model of the PMSM with faults, the fault decoupling is done for each fault presented in the system. From this fault, decoupling analysis is obtained subsystems with sensibility to a particular fault and without sensibility for the rest.

Subsystem sensitive to the control input fault f_a: For this subsystem, decoupling the output y_3 is not used to avoid the sensor fault effect. Considering the first two equations and the outputs y_1 and y_2 of the faulty system Eq. (21) a subsystem sensitive to the actuator fault is obtained, as follows:

$$\dot{x}_1 = -\frac{R}{L}x_1 + Px_3x_2 + \frac{1}{L}u_d + f_a, \tag{22}$$

$$\dot{x}_2 = -\frac{R}{L}x_2 - Px_3x_1 - \frac{P\phi}{L}x_3 + \frac{1}{L}u_q + f_a, \tag{23}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/L0 \end{bmatrix} \frac{\partial H}{\partial \mathbf{x}}, \tag{24}$$

solving Eq. (23) for x_3

$$x_3 = \frac{L}{P\phi + PLx_1}\left(-\frac{R}{L}x_2 - \dot{x}_2 + \frac{1}{L}u_q + f_a\right), \tag{25}$$

now replacing x_3 in Eq. (22)

$$\dot{x}_1 = -\frac{R}{L}x_1 + \frac{Lx_2}{\phi + Lx_1}\left(-\frac{R}{L}x_2 - \dot{x}_2 + \frac{1}{L}u_q + f_a\right) + \frac{1}{L}u_d + f_a, \tag{26}$$

multiplying Eq. (26) by $\left(\frac{\phi}{L} + x_1\right)$ and solving for $\frac{\phi}{L}\dot{x}_1 + x_1\dot{x}_1 + x_2\dot{x}_2 = \dot{v}_1$,

$$\begin{aligned} \dot{v}_1 = {} & \frac{R}{L}v_1 - \frac{R}{2L}(x_1^2 + x_2^2) + \frac{u_q}{L}x_2 + \frac{u_d}{L}\left(\frac{\phi}{L} + x_1\right) \\ & + f_a\left(x_2 + \frac{\phi}{L} + x_1\right), \end{aligned} \tag{27}$$

$$y_{v1} = v_1, \tag{28}$$

Eqs. (27) and (28) are the subsystem 1 with sensitivity to the control input fault f_a, where x_1 and x_2 are quantities available in measurable outputs y_1 and y_2, respectively.

Subsystem sensitive to the load torque fault $\Delta\tau_L$: once more the output y_3 is not used to avoid sensitivity to the sensor fault. Subtracting Eq. (22) to Eq. (23)

$$(\dot{x}_1 - \dot{x}_2) = -\frac{R}{L}(x_1 - x_2) + Px_3(x_1 + x_2) + \frac{1}{L}(u_d - u_q) + \frac{P\phi}{L}x_3, \tag{29}$$

if a new state $\dot{v}_2 = \dot{x}_1 - \dot{x}_2$ is defined, Eq. (29) becomes Eq. (30), this equation and the third equation of Eq. (21) define the subsystem 2,

$$\dot{v}_2 = -\frac{R}{L}v_2 + Px_3(v_2 + 2x_2) + \frac{1}{L}(u_d - u_q) + \frac{P\phi}{L}x_3, \tag{30}$$

$$\dot{x}_3 = \frac{3P\phi}{2J_m}x_2 - \frac{B}{J_m}x_3 - \frac{\tau_L}{J_m} + \Delta\tau_L, \tag{31}$$

$$\mathbf{y_{v2}} = \begin{bmatrix} v_2 \\ x_3 \end{bmatrix}, \tag{32}$$

where x_2 is available in the measurable output y_2.

Subsystem sensitive to the sensor fault $\Delta\omega$*:* since this subsystem must be sensitive to the sensor fault, the output y_3 is used. Using the transformed state $\dot{v}_3 = \dot{x}_1 - \dot{x}_2$, the subsystem 3 is obtained with sensibility to the sensor fault:

$$\dot{v}_3 = (Px_3 - \frac{R}{L})v_3 + 2Px_2x_3 + \frac{P\phi}{L}x_3 + \frac{1}{L}(u_d - u_q), \tag{33}$$

$$y_{v3} = v_3, \tag{34}$$

where x_3 and x_2 are quantities available in the measurable outputs y_1 and y_2, respectively.

Once decoupled subsystems were obtained, for the residual generator design an observer for each one for each of the decoupled subsystem is designed.

For decoupled subsystems sensitive to f_a and $\Delta\omega$, the observer design using the proposed approach in Ref. [26] coincides with a Luenberger observer [29, 30], but, however, this does not apply for decoupled subsystems sensitive to $\Delta\tau_L$.

The observer design and the residual generator for the decoupled subsystem sensitive to $\Delta\tau_L$ are presented. The decoupled subsystem sensitive to $\Delta\tau_L$ can be expressed as follows:

$$\begin{aligned} \dot{v}_2 &= -\frac{R}{L}v_2 + Px_3(v_2 + 2y_2) + \frac{P\phi}{L}x_3 + \frac{1}{L}(u_d - u_q), \\ \dot{x}_3 &= \frac{3P\phi}{2J_m}y_2 - \frac{B}{J_m}x_3 - \frac{\tau_L}{J_m} + \Delta\tau_L, \\ \mathbf{y_{v2}} &= \begin{bmatrix} v_2 \\ x_3 \end{bmatrix}. \end{aligned} \tag{35}$$

Which can be written in the form Eq. (9) with

$$H(x) = \frac{1}{2}\frac{R}{L}v_2^2 + \frac{1}{2}\frac{B}{J_m}x_3^2, \tag{36}$$

http://dx.doi.org/10.5772/68084

where $\mathbf{x_a} = [v_2\ x_3]^T$ whose gradient vector is defined as follows

$$\frac{\partial H(\mathbf{x_a})}{\partial \mathbf{x_a}} = \left[\frac{R}{L}v_2,\ \frac{B}{J_m}x_3\right]^T \Rightarrow \mathbf{M} = diag\left[\frac{R}{L},\ \frac{B}{J_m}\right], \tag{37}$$

and with

$$\mathbf{J}(\mathbf{x_a}) = \begin{bmatrix} 0 & \frac{J_m P\phi}{2BL} \\ -\frac{J_m P\phi}{2BL} & 0 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} -1 & \frac{J_m P\phi}{2BL} \\ \frac{J_m P\phi}{2BL} & -1 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_d - u_q \\ \Delta\tau_L \end{bmatrix},$$

$$\mathbf{F}(\mathbf{x_a}) = \begin{bmatrix} Px_3(v_2 + 2y_2) \\ \frac{3P\phi}{2J_m}y_2 - \frac{\tau_L}{J_m} \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \frac{1}{L} & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} \frac{L}{R} & \frac{J_m}{B} \end{bmatrix}.$$

For this case, the pair $(\mathbf{C}, \mathbf{S})$ is observable, and thus, there exists a matrix $\mathbf{K}$ that satisfies both the requirements of Theorem 1 for the observer design and the requirements of Theorem 2 for the residual generator design, and thus the observer is as follows

$$\begin{aligned} \dot{\hat{v}}_2 &= \frac{P\phi}{L}\hat{x}_3 - \frac{R}{L}\hat{v}_2 + P\hat{x}_3(\hat{v}_2 + 2y_2) + \frac{1}{L}(u_d - u_q) + L_2(v_2 - \hat{v}_2), \\ \dot{\hat{x}}_3 &= -\frac{B}{J_m}\hat{x}_3 - \frac{\tau_L}{J_m} + \frac{3P\phi}{2J_m}y_2 + L_3(x_3 - \hat{x}_3), \\ \hat{\mathbf{y}}_{\mathbf{v2}} &= \begin{bmatrix} \hat{v}_2 \\ \hat{x}_3 \end{bmatrix}. \end{aligned} \tag{38}$$

and the directional residual generator for the decoupled subsystem sensitive to $\Delta\tau_L$ is given by

$$r_2 = v_2 - \hat{v_2}. \tag{39}$$

For decoupled subsystem sensitive to control input f_a Eqs. (27) and (28), the observer and its directional residual generator are as follows:

$$\begin{aligned} \dot{\hat{v}}_1 = &-\frac{R}{L}v_1 - \frac{R}{2L}(y_2^2 + y_1^2) + u_d\left(\frac{y_1}{L} + \frac{\phi}{L^2}\right) \\ &+ \frac{u_q y_2}{L} + L_1(v_1 - \hat{v}_1), \end{aligned} \tag{40}$$

$$\hat{y}_{v1} = \hat{v}_1, \tag{41}$$

$$r_1 = v_1 - \hat{v}_1. \tag{42}$$

Finally, for the decoupled subsystem sensitive to the sensor fault $\Delta\omega$ Eq. (33), the observer and its directional residual generator are as follows:

$$\dot{\hat{v}}_3 = -\frac{R}{L}v_3 + Py_3(y_1 + y_2) + \frac{P\phi}{L}y_3 + \frac{1}{L}(u_d - u_q) + L_4(v_3 - v_3), \tag{43}$$

$$\hat{y_{v3}} = \hat{v_3}, \tag{44}$$

$$r_3 = v_3 - \hat{v_3}. \tag{45}$$

To summarize the fault sensitivity results of each residual (associated to each subsystem), see **Table 1**.

Where the fault affecting the residual is indicated with √, and the symbol ∅ means that there is no connection between the fault and the corresponding residual.

As can be appreciated from **Table 1**, there is a one-to-one relationship between faults and residuals so that perfect decoupling has been attached. One nice thing of perfect decoupling is that the occurrence of faults can be detected and isolated without problems.

The following results were obtained by computer simulation. **Table 2** shows the considered faults. About 10% of the nominal value of each variable is the fault magnitude considered, where 34 . 62 is the nominal value of the control input on the stationary reference frame $dq0$, 100 rad/sec is the nominal value of the angular speed, and 1 . 4 Nm is the nominal value of the load torque.

Residual (subsystem)	**Fault**		
	f_1 **Actuator fault (f_a)**	f_2 **Sensor fault ($\Delta\omega$)**	f_3 **Change of charge ($\Delta\tau_L$)**
1	√	∅	∅
2	∅	√	∅
3	∅	∅	√

Table 1. Fault incidence table.

Case	**Fault**	**Fault interval** (***sec***.)	**Magnitude**
1	0	0	0
2	f_a	[2, 2 . 5]	3 . 462
3	$\Delta\omega$	[3, 3 . 5]	10
4	$\Delta\tau_L$	[4, 4 . 5]	0 . 14
5	f_a, $\Delta\omega$, $\Delta\tau_L$	[3, 3 . 5], [4, 4 . 5], [2, 2 . 5]	3 . 462, 10, 0.14

Table 2. Fault cases.

Table 3 shows the PMSM parameters, which were taken from Ref. [22].

Figure 3 shows the evolution of PMSM states in the time (the time scale is given in seconds), that is, the current in the direct axis $i_d(t)$, the current in the quadrature axis $i_q(t)$ and the angular velocity $\omega(t)$ in nominal conditions (without faults). Actually, this figure represents the response of the PMSM with nominal parameters.

The evolution of the residuals when an actuator fault f_a occurs is depicted in **Figure 4**, where both residuals 2 and 3 are equal to zero since these are insensitive to the fault f_a, while residual 1 is different from zero indicating the sensitivity to control input fault f_a. Note that the magnitude of the two first residuals represents deviations between nominal and measurement currents with respect to the time (time is given in seconds). The third residual represents the deviation of the nominal and measured angular velocity of the PMSM rotor.

Figure 5 shows the residuals evolution when the sensor fault occurs, where residuals 1 and 3 are zero at all time due to its insensitivity to this fault, while residual 2 differs from zero due to its sensitiveness to this fault.

Figure 6 shows the residuals evolution when the load torque fault occurs, where residuals 1 and 2 are null and residual 3 is different from zero, indicating the sensitivity to load torque fault. Note that the fault magnitude is of 0.14; however, the residual becomes a value around 3, that is, the effect of the fault is not directly the magnitude of this. Extra work is required in the design of the observer-based residual in order to get at the residual a more approximated value of the fault magnitude.

Figure 7 shows the case when all three faults occur, even if not at the same time. The effect of the faults is manifested in the correct residual. It means that the problem of fault isolation in multiple faults can be carried out effectively.

As a final note, it can be appreciated a minimum transient at the beginning of all residuals signals, and this transient does not affect the fault detection and isolation process.

Parameter	Numerical value
Resistance (R)	1.6 Ω
Rotor inertia (J_m)	76.5×10^{-6}kg m^2
Viscous friction coefficient (B)	4×10^{-6}Nm/rev/min
Flux linkage (Φ)	0.29Nm/A
Inductance (L)	9.4H
Load torque (τ_L)	1.4Nm
Pole pairs (P)	1

Table 3. PMSM parameters.

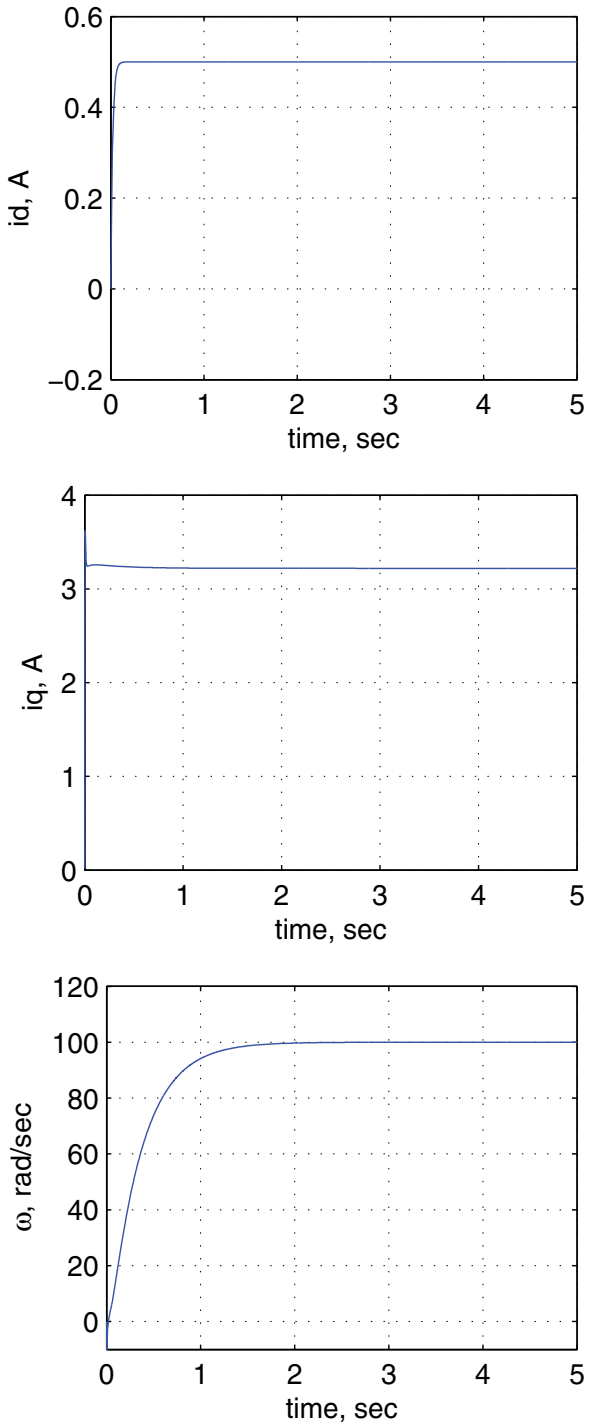


Figure 3. Nominal states.

http://dx.doi.org/10.5772/68084

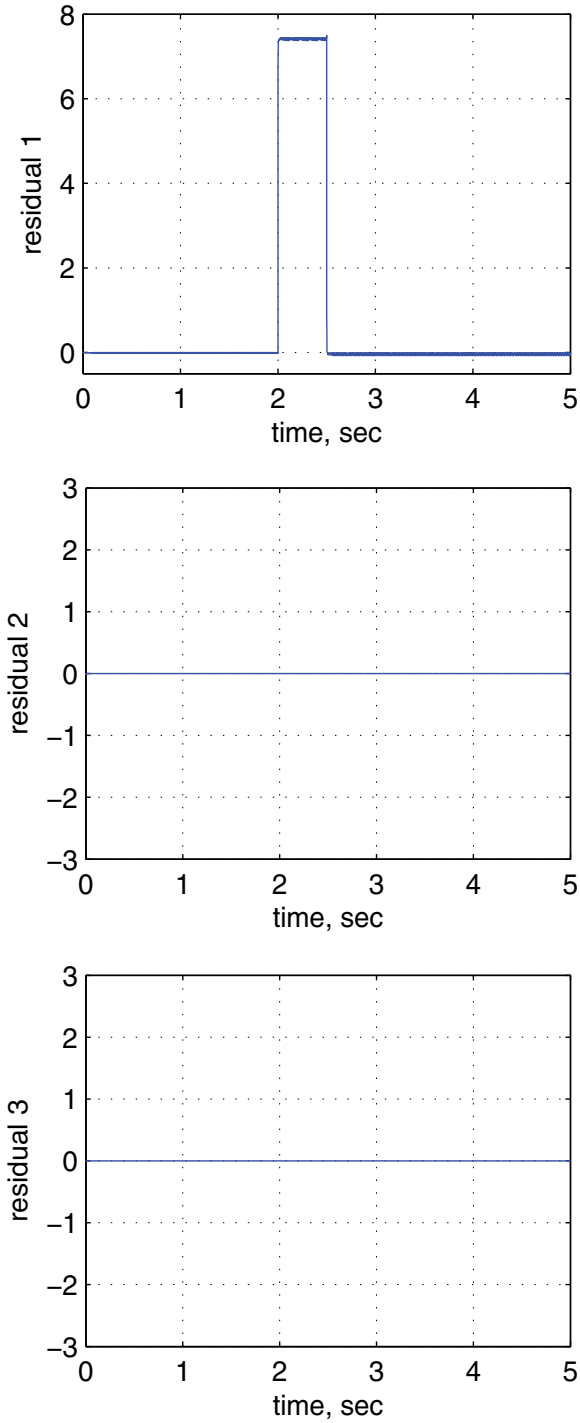


Figure 4. Residual sensitive to the actuator fault f_a.

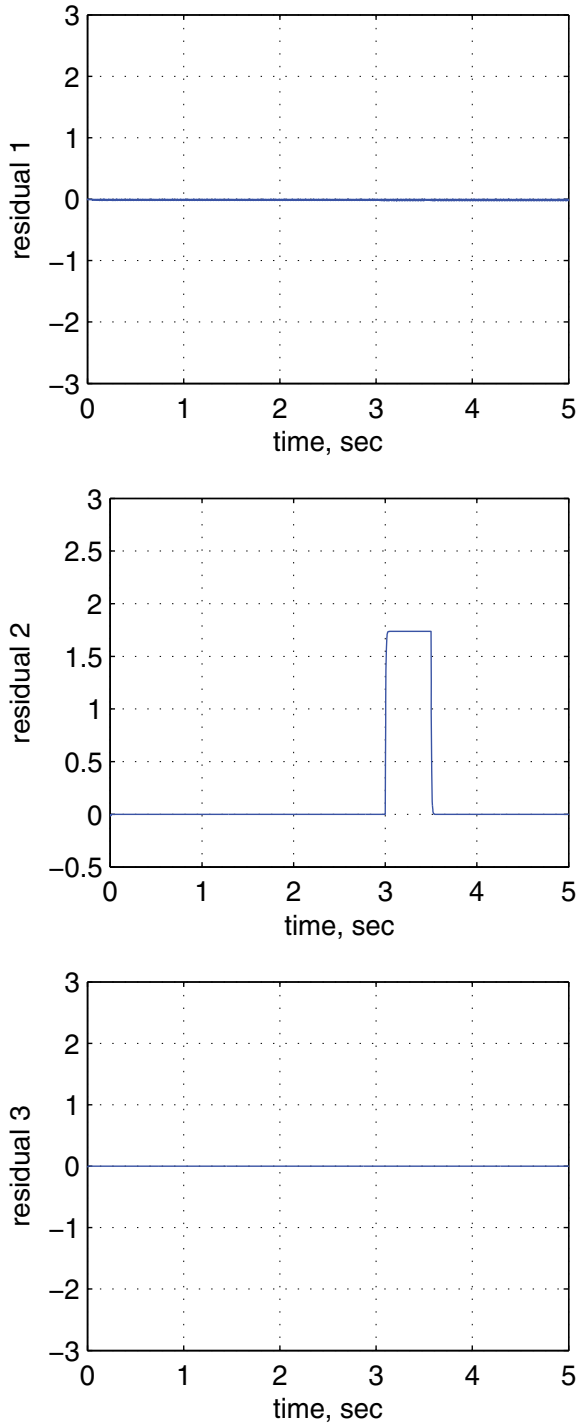


Figure 5. Residual sensitive to sensor fault $\Delta\omega$.

http://dx.doi.org/10.5772/68084

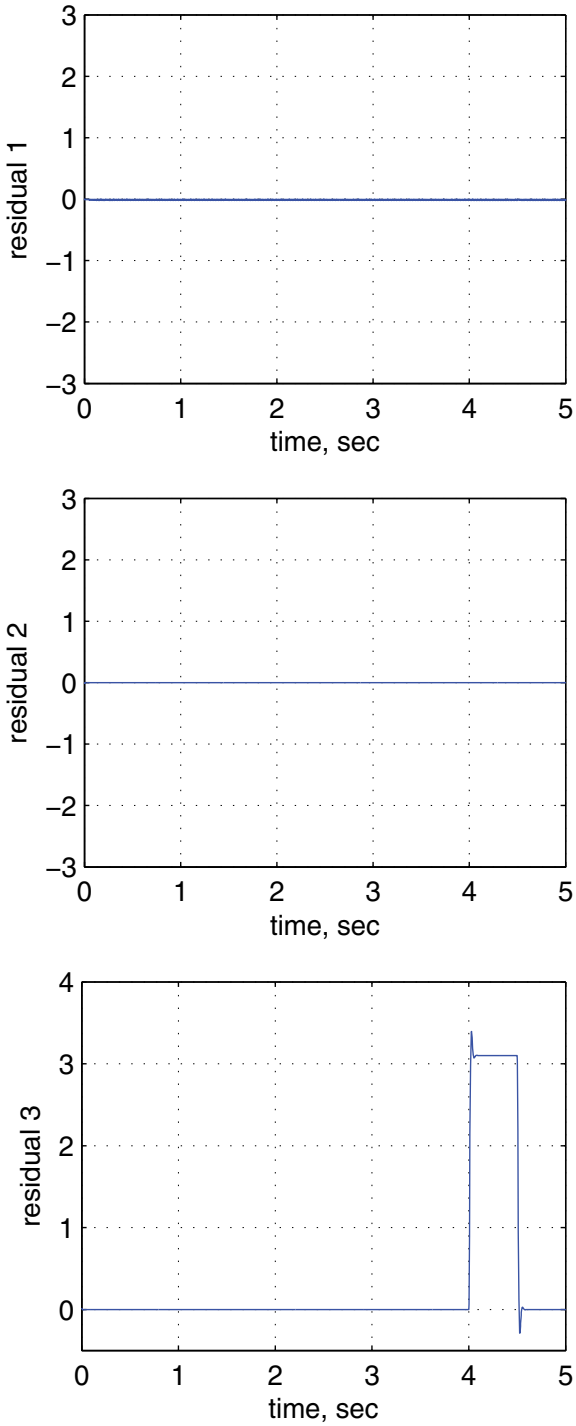


Figure 6. Residual sensitive to ΔT_L.

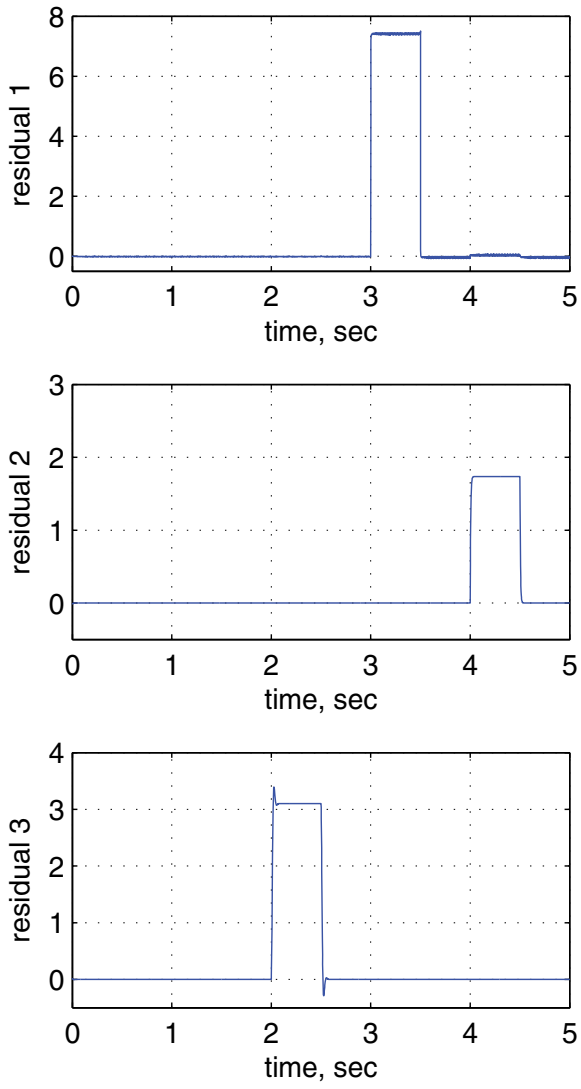


Figure 7. Residuals occurring simultaneously at f_a, Δ_ω and ΔT_L.

5. Concluding remarks

Fault diagnosis for a wide class of nonlinear systems, the class of systems that admit a Hamiltonian representation, has been considered. An observer-based solution with weak existence conditions for the fault diagnosis has been proposed, and this approach allows the detection and isolation of additive faults.

The proposed procedure follows the traditional way, namely: First, a decoupling methodology is applied to systems with Hamiltonian representation in order to obtain subsystems that preserve the Hamiltonian structure. Observer-based residual generators are designed for each subsystem so that each residual generator is sensible to a fault (or to a specific group of faults). The residual has the property of remain close to zero (or under a threshold value) if no fault is present in the system and non zero (or greater than a threshold value) when a fault affects the system. The proposed approach solves the fault isolation problem, and it permits a systematic design of the required residual generators. In contrast with other methodologies, for systems with Hamiltonian representation, an easy way to design an observer has been introduced. In addition, a wide set of nonlinear systems can be represented in the Hamiltonian structure, making the proposed solution widely applicable.

The proposed methodology has been applied to a synchronous machine, showing that, using the proposed approach, it is possible to detect and isolate additive faults in scenarios such as a fault in the control input, a change in the load torque as well as a fault in the angular velocity sensor.

Future work includes the study of multiplicative fault type.

Author details

Luis Humberto Rodriguez-Alfaro, Efrain Alcorta-Garcia*, Cornelio Posadas-Castillo and David Alejandro Diaz-Romero

*Address all correspondence to: efrain.alcortagr@uanl.edu.mx

Universidad Autónoma de Nuevo León México, San Nicolás de los Garza, Mexico

References

[1] Blanke M, Kinnaert M, Lunze J, Staroswiecki M. Diagnosis and fault-tolerant control. 2nd ed. Berlin, Germany: Springer Berlin Heidelberg; 2006. doi:10.1007/978-3-540-35653-0

[2] Chen J, Patton RJ. Robust model based fault diagnosis for dynamic systems. Kluwer Academic Publishers Group; 1999. doi:10.1007/978-1-4615-5149-2

[3] Ding SX. Model-based fault diagnosis techniques. Springer; 2008. doi:10.1007/978-3-540-76304-8

[4] Frank PM. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey. Automatica. 1990;26:459–474. doi:10.1016/0005-1098(90)90018-D

[5] Isermann R. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. 1st ed. Springer; 2006. doi:10.1007/3-540-30368-5

[6] Zhang K, Jiang B, Shi P. Fast fault estimation and accommodation for dynamical systems. IET Control Theory and Applications. 2009;3(2):189–199. doi:10.1049/iet-cta:20070283

[7] Alcorta Garcia E, Frank PM. Deterministic nonlinear observer based approaches to fault diagnosis: a survey. Control Engineering Practice. 1997;5(5):663–670. doi:10.1016/S0967-0661(97)00048-8

[8] Jiang B, Staroswiecki M, Cocquempot V. Fault estimation in nonlinear uncertain systems using robust sliding mode observers. IEE Control Theory and Application. 2004;151:29–37. doi:10.1049/ip-cta:20040074

[9] Persis CD, Isidori A. A geometric approach to nonlinear fault detection and isolation. IEEE Transactions on Automatic Control. 2001;46(6):853–865. doi:10.1109/9.928586

[10] Chen W, Saif M. Unknown input observer design for a class of nonlinear systems: an LMI approach. In: Proceedings of the American Control Conference. Mineapolis, MN; 2006. pp. 834–838.

[11] Pertew AM, Marquez HJ, Zhao Q. Design of unknown input observers for Lipschitz nonlinear systems. In: Proceedings of the American Control Conference. Portlend; 2005. pp. 4198–4203.

[12] Liu X, Yuan S. On designing H_∞ fault estimator for switched nonlinear systems of neutral type. Communications in Nonlinear Science and Numerical Simulation. 2011;6 (2011):4379–4389. doi:10.1016/j.cnsns.2011.03.017

[13] Jiang B, Wang JL. Actuator fault diagnosis for a class of bilinear systems with uncertainty. Journal of the Franklin Institute. 2002;339:361–374. doi:10.1016/S0016-0032(01)00051-5

[14] Xu J, Lum KY, Xie L, Loh AP. Fault detection and isolation of nonlinear systems: an unknown input observer approach with sum of squares techniques. Journal of Dynamic Systems, Measurement, and Control. 2012;134:041005-041005-7. doi:10.1115/1.4006074

[15] Bokor J, Szabó. Fault detection and isolation in nonlinear systems. Annual Reviews in Control. 2009;33:113–123. doi:10.1016/j.arcontrol.2009.09.001

[16] Lopes dos Santos P, AzevedoPerdicoilis TP, Jank G, Ramos JA, Martins de Carvalho JL. Leakage detection and location in gas pipelines through an LPV identification approach. Communications in Nonlinear Science and Numerical Simulation. 2011;16(2011):4657–4665. doi:10.1016/j.cnsns.2011.03.029

[17] Márton L, Ossmann D. Energetic approach for control surface disconnection fault detection in hydraulic aircraft actuators. In: Preprints of the 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS). Mexico; 2012. pp. 1149–1154.

[18] Samadani M, Kwuimy CAK, Nataraj C. Model-based fault diagnostics of nonlinear systems using the features of the phase space response. Communications in Nonlinear Science and Numerical Simulation. 2014;20(2015):583–593. doi:10.1016/j.cnsns.2014.06.010

[19] Rios H, Edwards C, Davila J, Fridman L. Fault detection and isolation for nonlinear systems via HOSM multiple-observer. In: Preprints of the 8th IFAC Symposium on Fault

Detection, Supervision and Safety of Technical Processes (SAFEPROCESS). Mexico; 2012. pp. 534–539.

[20] Ortega R, Espinoza-Pérez G, Astolfi A. Passivity based control of AC drives: theory for the user and examples. International Journal of Control. 2013;86:625–635. doi:10.1080/00207179.2012.753643

[21] van der Schaft A. L_2 gain and passivity techniques in nonlinear control. Springer; 2000. doi:10.1007/978-1-4471-0507-7

[22] Zhou J, Wang Y. Adaptive backstepping speed controller design for a permanent magnet synchronous motor. IEE Proceedings - Electric Power Applications. 2002;149(2):165–172. doi:10.1049/ip-epa:20020187

[23] Sira-Ramierez H, Cruz-Hernandez C. Synchronization of chaotic systems: a generelazed Hamiltonian systems approach. International Journal of Bifurcation and Chaos. 2001;11(5):1381–1395.

[24] Seliger R, Frank PM. Fault diagnosis by disturbance decoupled nonlinear observers. In: Proceedings of the CDC '91, Brighton, England; 1991. pp. 2248–2253.

[25] Alcorta Garcia E. ModelgestützeResiduengenerierungfür die Diagnose von additiven and multiplicativenFehler in dynamischenSystemen (in german) PhD [Thesis]. Gerhard-Mercator University of Duisburg (Duisburg-Essen Universität), Germany. Duesseldorf, Germany; 1999.

[26] Rodriguez-Alfaro LH, Alcorta-Garcia E, Lara D, Romero G. A Hamiltonian appraoch to fault isolation in a planar vertical take off and landing aircraft model. International Journal of Applied Mathematics and Computer Science. 2015;25(1):65–76. doi:10.1515/amcs-2015-0005

[27] Krause PC, Wasynczuk O, Sudhoff SD. Analysis of electric machinery and drive systems. John Wiley and Sons, Inc. Publication; 2002. doi:10.1109/9780470544167

[28] Campos-Delgado DU, Espinoza Trejo DR, Palacios E. Fault-tolerant control in variable speed drives: a survey. IET Electric Power Applications. 2008;2(2):121–134. doi:10.1049/iet-epa:20070203

[29] Luenberger DG. Observing the state of a linear system. IEEE Transactions on Military Electronics. 1964;8(2):74–80. doi:10.1109/TME.1964.4323124

[30] D'Azzo J, Houpis CH. Linear control system analysis and design with matlab. 5th ed. Marcel Dekker; 2003. doi:10.1201/9780203911426

Chapter 10

Process Monitoring Using Data-Based Fault Detection Techniques: Comparative Studies

Mohammed Ziyan Sheriff, Chiranjivi Botre,
Majdi Mansouri, Hazem Nounou,
Mohamed Nounou and Mohammad Nazmul Karim

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67347

Abstract

Data based monitoring methods are often utilized to carry out fault detection (FD) when process models may not necessarily be available. The partial least square (PLS) and principle component analysis (PCA) are two basic types of multivariate FD methods, however, both of them can only be used to monitor linear processes. Among these extended data based methods, the kernel PCA (KPCA) and kernel PLS (KPLS) are the most well-known and widely adopted. KPCA and KPLS models have several advantages, since, they do not require nonlinear optimization, and only the solution of an eigenvalue problem is required. Also, they provide a better understanding of what kind of nonlinear features are extracted: the number of the principal components (PCs) in a feature space is fixed a priori by selecting the appropriate kernel function. Therefore, the objective of this work is to use KPCA and KPLS techniques to monitor nonlinear data. The improved FD performance of KPCA and KPLS is illustrated through two simulated examples, one using synthetic data and the other using simulated continuously stirred tank reactor (CSTR) data. The results demonstrate that both KPCA and KPLS methods are able to provide better detection compared to the linear versions.

Keywords: principal component analysis, partial least squares, kernels, fault detection, process monitoring

1. Introduction

Process monitoring is an essential aspect of nearly all industrial processes, often required both to ensure safe operation and to maintain product quality. Process monitoring is generally carried out in two phases: detection and diagnosis. This chapter focuses only on the fault detection aspect. Fault detection methods can be categorized using a number of different methodologies. One popular method of categorization is into quantitative model-based methods, qualitative model-based methods, and data (process history)-based methods [1–3]. **Figure 1** illustrates a general schematic of fault detection phase.

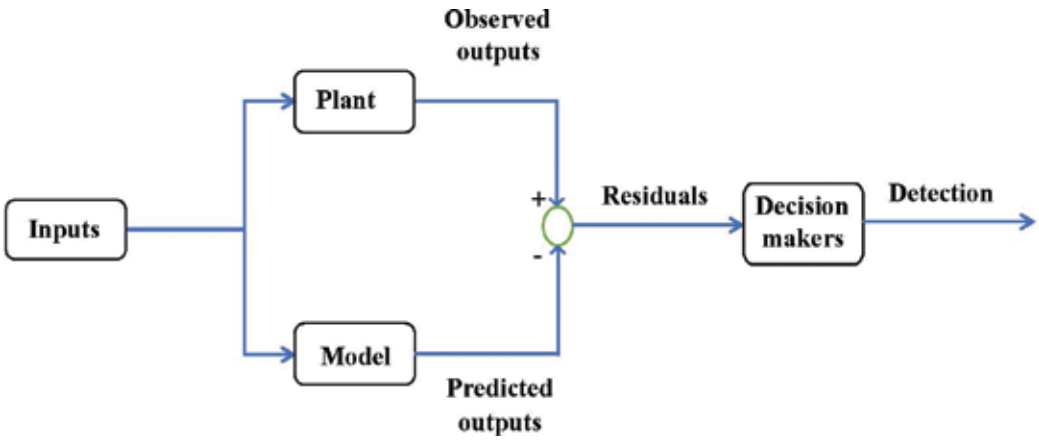


Figure 1. Schematic illustration of detection phase.

Quantitative model-based methods require knowledge of the process model, while qualitative model-based methods require expert knowledge of the given process. Hence, data-based methods are often used as they require neither prior knowledge of the process model nor expert knowledge of the process [4].

Data-based monitoring methods can be further classified into input model-based methods and input-output model-based methods. Input model-based methods only require the data matrix of the input process variables, while input-output model-based methods require both the input and output data matrices in order to formulate a model and carry out fault detection [5]. Input model-based methods are sometimes utilized when the input-output models cannot be formed due to the high dimensionality and complexity of a system being monitored [6]. However, input-output model-based methods do have the added advantage of being able to detect faults in both the process and the variables [5].

Principal component analysis (PCA) is a widely used input model-based method that has been used for monitoring a number of processes including air quality [7], water treatment [8], and semiconductor manufacturing [9]. On the other hand, partial least squares (PLS) are an input-output model-based method that has been applied in chemical processes to monitor online measurement variables and also to monitor and predict the output quality variable [10]. PLS

http://dx.doi.org/10.5772/67347

has been applied for the monitoring of distillation columns, batch reactors [11], continuous polymerization processes [12], and other similar industrial processes, which are described by input-output models. However, both PCA and PLS are fault detection techniques that only work reasonably well with linear data. PCA and PLS have been extended to handle nonlinear data by utilizing kernels to transform the data to a higher dimensional space, where linear relationships between variables can be drawn. The extensions kernel principal component analysis (KPCA) and kernel partial least squares (KPLS) have both shown improved performance over the conventional PCA and PLS techniques when handling nonlinear data [5, 13]. T^2 and Q charts are commonly used as fault detection statistics. In the literature, it has been seen that T^2 test is less effective fault detection technique compared to Q statistic; this is because T^2 test can only represent variation of the data in the principle component and not in residue of the model [14].

In our previous works [5, 13, 15], we addressed the problem of fault detection using linear and nonlinear input models (PCA and kernel PCA) and input-output model (PLS and kernel PLS)-based generalized likelihood ratio test (GLRT), in which PCA, kernel PCA, PLS, and kernel PLS methods are used for modeling and the univariate GLRT chart is used for fault detection. In the current work, we propose to use the PCA, kernel PCA, PLS, and kernel PLS methods for multivariate fault detection through their multivariate charts Q and T^2. The fault detection performance is evaluated using two examples, one using simulated synthetic data and the other utilizing a simulated continuous stirred tank reactor (CSTR) model.

The remainder of this chapter is organized as follows. Section 1 introduces linear PCA and PLS, along with the fault detection indices used for these methods. Section 2 then describes the idea of using kernels for nonlinear transformation of data, along with the kernel fault detection extensions: KPCA and KPLS. In Section 3, two illustrative examples are presented, one using simulated synthetic data and the other utilizing a simulated continuous stirred tank reactor. At the end, the conclusions are presented in Section 4.

2. Conventional linear fault detection methods

Before constructing either the PCA or PLS models, data are generally preprocessed to ensure that all process variables in the data matrix are scaled to zero mean and unit variance. This step is essential as different process variables are usually measured with varying standard deviations and means and often using different units.

2.1. Principal component analysis (PCA)

Consider the following input data matrix, $X \in R^{n \times m}$, where m and n represent the number of process variables and the number of observations, respectively. After preprocessing the data, single value decomposition (SVD) can be utilized to express the input data matrix as follows:

$$X = TP^T \tag{1}$$

where $T = [t_1, t_2, t_3 \ldots t_m] \in R^{n \times m}$ is a matrix of the transformed variables, where each column represents the score vectors or the transformed variables, and $P = [p_1, p_2, p_3 \ldots p_m] \in R^{m \times m}$ is a matrix of the orthogonal vectors, where each column is also known as loading vectors, and these are eigenvectors that are associated with the covariance matrix of the input data matrix X. The covariance matrix can be computed as follows [13]:

$$\sum = \frac{1}{n-1} X^T X = P \Lambda P^T \text{ with } PP^T = P^T P = I_m \tag{2}$$

where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is a diagonal matrix that contains the eigenvalues that are related to the m principal components ($\lambda_1 > \lambda_2 > \ldots > \lambda_m$), and I_m is the identity matrix [16]. It should be noted that the model built by PCA uses the same number of principal components as the original number of process variables in the input data matrix (m). However, since many industrial processes may contain process variables that are highly correlated, a smaller number of principal components can be utilized to capture the variation in the process data [6]. The quality of the model built by PCA is dictated by the number of principal components obtained. Overestimating the number could introduce noise that may mask important features in the data, while underestimating the number could decrease the prediction ability of the model [17].

Therefore, selection of the number of principal components is vital, and several methods have been developed for this purpose. A few popular techniques are cumulative percent variance (CPV) [13], scree plot and profile likelihood [18], and cross validation [19]. CPV is commonly utilized due to its computational simplicity and because it provides a good estimate of the number of principal components that need to be retained for most practical applications. CPV can be computed as follows [13]:

$$CPV(l) = \frac{\sum_{i=1}^{l} \lambda_i}{trace(\Sigma)} \times 100 \tag{3}$$

CPV is used to select the smallest number of principal components that represents a certain percentage of the total variance (e.g., 99%). Once the number of principal components to retain is determined, the input data matrix can then be expressed as [13]:

$$X = TP = \left[\widehat{T}\ \tilde{T}\right]\left[\widehat{P}\ \tilde{P}\right]^T \tag{4}$$

where $\widehat{T} \in R^{n \times l}$ and $\tilde{T} = \in R^{n \times m-l}$ represent the matrices containing the l retained principal components and the ignored $(m - l)$ principal components, respectively. Likewise, the matrices that contain the l retained eigenvectors and the ignored $(m - l)$ eigenvectors are represented by $\widehat{P} \in R^{m \times l}$ and $\tilde{P} \in R^{m \times m-l}$, respectively.

After expansion X can be expressed as [13]

$$X = \widehat{T}\widehat{P}^T + \tilde{P}\tilde{T}^T = \overbrace{X\widehat{P}\widehat{P}^T}^{\widehat{X}} + \overbrace{X\left(I_m - \widehat{P}\widehat{P}^T\right)}^{E} \tag{5}$$

where matrix $\widehat{X}$ is the modeled variation of X computed utilizing only the l retained principal components, while matrix E represents the residual space formed by variations that correspond to process noise.

The PCA model can be illustrated as shown in **Figure 2**.

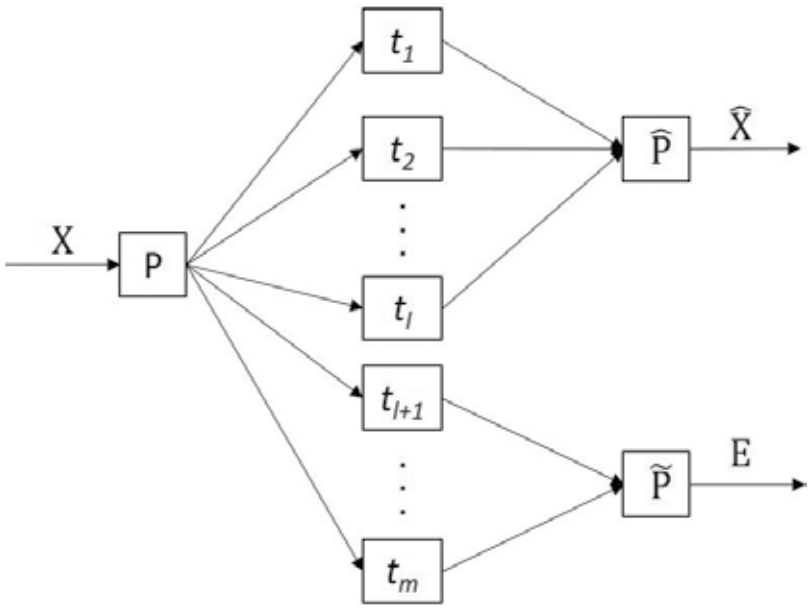


Figure 2. Schematic illustration of PCA.

2.2. Partial least squares (PLS)

PLS is a popular input-output technique used for modeling, regression and as a classification tool, which has been extended to fault detection purpose [20]. PLS includes process variables ($X \in R^{n \times m}$) and the quality variables ($Y \in R^{n \times p}$) with a linear relationship between input and output score vectors. Nonlinear iterative partial least square (NIPALS) algorithm developed by Word et al. is used to compute score matrices and loading vectors [21]:

$$\begin{aligned} X &= TP^T + E = \sum_{i=1}^{M} t_i p_i + E \\ Y &= UQ^T + F = \sum_{j=1}^{M} u_j q_j^t + F \end{aligned} \tag{6}$$

where $E \in R^{n \times m}$ and $F \in R^{n \times p}$ are the PLS model residues; $T \in R^{n \times M}$ and $U \in R^{n \times M}$ are the orthonormal input and output score matrix, respectively; P and Q are the loading vectors of the input (X) and output (Y) matrices, respectively; m and n are the number of process variables and observations in input (X) matrix; p is the number of quality variables in output (Y) matrix; and M is the total number of latent variables extracted. NIPALS method is shown in

Algorithm 1; X and Y matrixes are first standardized by mean centering and unit variance. NIPALS algorithm is initialized by assigning one of the columns of output matrix (Y) as output score vector (u); at each iteration t, u, p, and q are computed and stored; M latent variables are extracted.

Another modification to NIPALS algorithm has been published in the literature [22]. In other work, different variations of PLS technique have been stated. Qin et al. have presented recursive PLS model [23], where PLS model is updated with new training data set; MacGregor et al. [24] have developed multiblock PLS model to monitor subsection of process variables. While for process monitoring of batch processes, PLS has been extended to multiway PLS technique [25] to incorporate past batches in training data set.

PLS being an input-output type model can also be used as a regression tool, to predict quality variable (Y) from online measurement variable (X). From PLS, model input and output matrices are related by

$$Y = BX + G \tag{7}$$

The regression coefficient B is computed as shown in Eq. (8):

$$B = W\left(P^T W\right)^{-1} C^T \tag{8}$$

Substituting weights, loading vector P and constant C from Algorithm 1:

$$B = X^T U\left(T^T X X^T Y\right)^{-1} T^T Y \tag{9}$$

From Eqs. (7) and (9), the output matrix (Y) is predicted as

$$Y_{new} = X_{new} X^T U\left(T^T X X^T U\right)^{-1} T^T Y + G \tag{10}$$

Algorithm 1: Modified NIPALS algorithm

1. Initialized output score: $u = y_i$. (11)

2. Weights regressed on X: $w = u^T X / u^T u$. (12)

3. Normalizing weight: $w = w / \|w\|$. (13)

4. Weights R: $r_1 = w_1$ (14)

$$r_j = \prod_{i=1}^{j-1} \left(I_{m\times m} - w_i p_j^T\right) w_j, j > 1.$$

5. Input Score vector: $t = Xr / r^T r$. (15)

6. Input loading vector: $p = Xt^T / t^T t$. (16)

7. Output loading vector: $q = Yt^T / t^T t$. (17)

8. Output score vector: $u = Yq/q^T q$. (18)

9. Normalizing weights, loading vectors and scores:

$$p = \frac{p}{norm(p)}, w = w \times norm(p), t = t \times norm(p). \tag{19}$$

10. Input and output matrices are deflated:

$$\begin{aligned} X &= X - tp^T \\ Y &= Y - tq^T \end{aligned} \tag{20}$$

11. Store latent score vectors in T and U, loading vectors in P and Q

12. Repeat steps 2 to 11 until M latent variables are computed

2.3. Fault detection indices

Different fault detection indices can be used for the linear PCA and PLS techniques. The two most popular indices are the T^2 and Q statistics. T^2 measures the variation of the model, while the Q statistic measures the variation in the residual space, and these statistics will be described next.

2.3.1. T^2 *statistic*

The T^2 statistic measures the variation in the principal components at different time samples and is defined as follows [26, 27]:

$$T^2 = X^{\mathbf{T}} \widehat{\mathbf{P}} \widehat{\Lambda} \widehat{\mathbf{P}}^{\mathbf{T}} X \tag{21}$$

where $\widehat{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_l)$ is the diagonal matrix that contains the eigenvalues that are associated with the retained principal components. For testing data, a fault is declared when the T^2 value exceeds the value of the threshold as follows:

$$T^2 \geq T_\alpha^2 = \frac{(n^2 - 1)l}{n(n - l)} F(l, n - l) \tag{22}$$

where α is the level of significance, generally assigned a value between 90 and 99%, and F (a, $n - a$) is the critical value of the Fisher-Snedecor distribution with n and $n-a$ degrees of freedom.

2.3.2. Q statistic

The Q statistic measures the projection of the data on to the residual subspace and allows the user to measure how well the data fit the PCA model. The Q statistic is defined as follows [16]:

$$Q = \|\tilde{X}\| = \left\| \left(I - \widehat{P}\widehat{P}T \right) X \right\|^2 \tag{23}$$

For testing data, a fault is declared when the threshold value is violated as follows [16]:

$$Q \geq Q_\alpha = \varphi_1 \frac{h_0 c_\alpha \sqrt{2\varphi_2}}{\varphi_1} + 1 + \frac{\varphi_2 h_0 (h_0 - 1)}{\varphi_1^2} \tag{24}$$

where $\varphi_i = \sum_{j=l+1}^{m} \lambda_j^i \;\; i = 1, 2, 3$, $h_0 = 1 - \frac{2\varphi_1 \varphi_3}{3\varphi_2^2}$, where c_α is the value obtained from the normal distribution of significance α.

3. Nonlinear fault detection methods using kernel transformations

A popular nonlinear version of PCA and PLS is the projection of nonlinear data to a high-dimensional feature space, where the linear fault detection method is applied in the features space, *F*. The authors in Ref. [28] used projection of *X* for PLS response surface modeling using the quadratic function as the mapping function:

$$\Phi : \chi = R^2 \rightarrow F = R^3 \tag{25}$$

However, it is difficult to know the accurate nonlinear transformation function for nonlinear data matrix to be linear in the feature space. According to Mercer's theorem, orthogonal semi-positive definite function can be used to map the data into the feature space instead of knowing the explicit nonlinear function. This nonlinear function is called the kernel function and is defined as the dot product of the mapped data in the feature space:

$$k(X_i, X_j) = \Phi(X_i)\Phi(X_j) \tag{26}$$

Thus, kernel-based multivariate methods can be defined as nonlinear fault detection methods in which the input data matrix is mapped into high-dimensional feature space and developed linear models can be applied in the feature space for fault detection purposes.

Commonly used kernel functions are given below [29]:

Radial basis function (RBF):

$$K(X, Y) = \exp\left(\frac{-\|X - Y\|^2}{c}\right) \tag{27}$$

Polynomial function:

$$K(X, Y) = \langle X, Y \rangle^d \tag{28}$$

Sigmoid function:

$$K(X, Y) = \tanh(\beta_0 \langle X, Y \rangle + \beta_1) \tag{29}$$

http://dx.doi.org/10.5772/67347

The next section describes the methodology of utilizing kernel transformations to extend linear PCA and PLS to the hyperdimensional space in order to carry out fault detection of nonlinear data.

3.1. Kernel principal component analysis (KPCA)

While PCA seeks to find the principal components by minimizing the data information loss in the input space, KPCA does this in the feature space (F). For KPCA learning using training data, $X_1, X_2, \ldots, X_n \in R^m$, nonlinear mapping gives Φ: $X \in \Re^m \rightarrow Z \in \Re^h$, where input data are extended into the hyperdimensional feature space, where the dimension can be very large and possibly infinite [30].

The covariance in the feature space can be computed as follows [31]:

$$\mathbf{C}^F = \frac{1}{n}\sum_{j=1}^{n} \Phi(X_j)\Phi(X_j)^T \tag{30}$$

Similar to PCA, the principal components in the feature space can be found by diagonalizing the covariance matrix. In order to diagonalize the covariance matrix, it would be necessary to solve the following eigenvalue problem in the feature space [31]:

$$\lambda \mathbf{v} = \mathbf{C}^F \mathbf{v} \tag{31}$$

where $\lambda \geq 0$ and represents the eigenvalues.

In order to solve the eigenvalue problem, the following equation is derived [32]:

$$n\lambda\alpha = K\alpha \tag{32}$$

where K and α are the $n \times n$ kernel matrix and eigenvectors, respectively.

For test vector X, the principal components (t) are extracted projecting $\Phi(X)$ onto the eigenvectors $\mathbf{v}_k$ in the feature space where $k = 1,\ldots,l$:

$$\mathbf{t}_k = \langle \mathbf{v}_k, \Phi(X) \rangle = \sum_{i=1}^{N} \alpha_i^k \langle \Phi(X_i), \Phi(X) \rangle \tag{33}$$

It is important to note that before carrying out KPCA, it is necessary to mean center the data in the high-dimensional space. This can be accomplished by replacing the kernel matrix K with the following [32]:

$$\mathbf{K} = \mathbf{K} - 1_n\mathbf{K} - \mathbf{K}1_n + 1_n\mathbf{K}1_n \tag{34}$$

where $1_n = \frac{1}{n}\begin{vmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{vmatrix} \in R^{n\times n}$.

3.1.1. T^2 *statistic for KPCA*

Variation in the KPCA model can be found using T^2 statistic, which is the sum of normalized squared scores, computed as follows [31]:

$$T^2 = [\mathbf{t}_1, \ldots, \mathbf{t}_l]\Lambda^{-1}[\mathbf{t}_1, \ldots, \mathbf{t}_l]^T \tag{35}$$

where $\mathbf{t}_k$ is obtained from Eq. (33).

The confidence limit is computed as follows [31]:

$$T^2_{l,n,\alpha} \sim \frac{l(n-1)}{n-l} F_{l,n-l,\alpha} \tag{36}$$

3.1.2. Q statistic for KPCA

In order to compute the Q statistic, the feature vector $\Phi(X)$ needs to be reconstructed. This is done by projecting $\mathbf{t}_k$ into the feature space using $\mathbf{v}_k$ as follows [31]:

$$\widehat{\Phi}_n(X) = \sum_{k=1}^{n} \mathbf{t_k v_k} \tag{37}$$

The Q statistic in the feature space can now be computed as [31]

$$Q = \left\| \Phi(X) - \widehat{\Phi}_l(X) \right\|^2 \tag{38}$$

The confidence limit of the Q statistic can then be computed using the following equation [31]:

$$Q_\alpha \sim g\chi^2_h \tag{39}$$

This limit is based on Box's equation, obtained by fitting the reference distribution obtained using training data, to a weighted distribution. Parameter g is the weight assigned to account for the magnitude of the Q statistic, and h represents the degree of freedom. Considering a and b the estimated mean and variance of the Q statistic, g and h are approximated using $g = b/2a$ and $h = 2a^2/b$.

3.2. Kernel partial least square (KPLS)

The KPLS methodology works by mapping the data matrices into the feature space and then applying the nonlinear partial least square algorithm and computing the loading and score vectors.

The mapped data points are given as

$$\begin{gathered} \Phi : \chi \to F \\ \chi \to \left(\sqrt{\lambda_1}\varphi_1(X), \sqrt{\lambda_2}\varphi_2(X), \ldots, \sqrt{\lambda_n}\varphi_n(X) \right) \end{gathered} \tag{40}$$

The kernel gram function can be used to map the data into the feature space instead explicitly using the nonlinear mapping function; this is called the kernel trick. Kernel gram function is defined as the dot product of the mapping function:

$$k(X_i, X_j) = \Phi(X_i)\Phi(X_j) \tag{41}$$

As with the KPCA algorithm, the kernel matrix has to be mean centered before applying the NIPALS algorithm using Eq. (34).

The input score matrix and weights are computed as

$$\begin{aligned} t &= \overline{\phi}(X)^T R \\ R &= \Phi^T U\left(T^T \overline{K} U\right)^{-1} \end{aligned} \tag{42}$$

Thus, the score matrix is given as [33]

$$t = K_{tt} Z \tag{43}$$

where $Z = U\left(T^T K U\right)^{-1}$.

Now, the relationship between the input and output score matrices can be derived by combing Eqs. (15), (17), and (18):

$$t = XX^T u \tag{44}$$

In the feature space, replace X by its image Φ:

$$t = \Phi\Phi^T u \tag{45}$$

Substituting the kernel gram function, $K = \Phi\Phi^T$, input and output scores are given by

$$\begin{aligned} t &= Ku \\ u &= Yt \end{aligned} \tag{46}$$

After every iteration, input kernel (K) and output matrix (Y) are deflated as

$$\Phi\Phi^T = \left(\Phi - tt^T\Phi\right)\left(\Phi - tt^T\Phi\right)^T \tag{47}$$

$\Phi\Phi^T$ dot product is replaced by kernel gram function K:

$$\begin{aligned} K &= K - tt^T K - Ktt^T + tt^T Ktt^T \\ Y &= Y - tt^T Y \end{aligned} \tag{48}$$

Let $\{X_i\}_{i=1}^n$ be the training data and $\left\{X_j\right\}_{j=1}^n$ be the testing data, $\Phi(X_i)$ is the mapped training data, and $\Phi(X_j)$ is the mapped testing data. Kernel functions for the testing data are given as

$$K_t = K(X, X_t) = \sum_{i=1}^{n} \sqrt{\lambda_i}\varphi_i(X)\sqrt{\lambda_i}\varphi_i(X_t) = (\Phi(X) \cdot \Phi(X_t)) = \Phi(X)^T\Phi(X_t)$$
$$K_{tt} = K(X_t, X_t) = \sum_{i=1}^{n} \sqrt{\lambda_i}\varphi_i(X_t)\sqrt{\lambda_i}\varphi_i(X_t) = (\Phi(X_t) \cdot \Phi(X_t)) = \Phi(X_t)^T\Phi(X_t) \tag{49}$$

KPLS algorithm can also be used to predict output matrix Y from input matrix X as

$$Y_t = \Phi_t B \tag{50}$$

Φ_t is mapped testing data in feature space from $\{X_j\}_{j=1}^{n}$, and B is the regression coefficient which is given as [34]

$$B = \Phi^T Y\left(T^T K U\right)^{-1} T^T Y \tag{51}$$

Thus combining Eqs. (50) and (51), we get predicted output quality matrix:

$$Y_t = K_t U\left(T^T K U\right)^{-1} T^T Y \tag{52}$$

Algorithm 2: Kernel partial least square (KPLS) algorithm

1. Compute Kernel matrix: K.
2. Kernel matrix is mean centered using Eq. (34).
3. For first iteration, initialized score matrix: $u = y_i$.
4. Calculate scores t and u using Eq. (46).
5. Deflate K and Y, using Eq. (48)
6. Score vectors t and u are stored in cumulative matrix T and U
7. Repeat steps 1 to 6 to extract M latent variables.

3.2.1. T^2 statistic for KPLS

The T^2 statistic for KPLS can be computed as

$$T_t^2 = t^T \Lambda^{-1} t \tag{53}$$

where $\Lambda = (n-1)^{-1} T^T T$ and the score matrix being orthonormal matrix $T^T T = I$, leading to $\Lambda = (n-1)^{-1} I$. The score matrix is $t = K_{tt} Z$; hence the T^2 statistic is given by [33]

$$T_t^2 = (n-1) K_{tt} Z Z^T K_{tt} \tag{54}$$

The threshold value for T^2 statistic is computed using the f-inverse distribution and is given by [35]

$$T_{\alpha}^{2} = g \cdot \text{finv}(\alpha, m, h) \tag{55}$$

where n and m are the total number of observations and variables in the input data matrix X, respectively, and

$$\begin{aligned} g &= \frac{m\left(n^2-1\right)}{n(n-m)} \\ h &= n-m \end{aligned} \tag{56}$$

3.2.2. Q statistic for KPLS

As with the other data-based models, the Q statistic computes the mean square error of the residue from the KPLS model:

$$\begin{aligned} Q &= \left\| \overline{\phi} - \tilde{\overline{\phi}} \right\|^2 \\ Q &= \overline{\phi}^T \overline{\phi} - 2\overline{\phi}^T \tilde{\overline{\phi}} + \tilde{\overline{\phi}}^T \tilde{\overline{\phi}} \end{aligned} \tag{57}$$

Substituting the kernel gram functions as the dot product of mapped points:

$$Q = K_{tt} - 2K_t^T KZt + t^T T^T KTt \tag{58}$$

where $Z = U\left(T^T KU\right)^{-1}$.

The threshold value for the Q statistic under the significance level of α [36] is given by

$$Q_{\alpha} = g\chi_{\alpha}^{2}(h) \tag{59}$$

where g and h are given by

$$\begin{aligned} g &= \frac{variance(Q)}{2 \times mean(Q)} \\ h &= \frac{2 \times (mean(Q))^2}{variance(Q)} \end{aligned} \tag{60}$$

A fault is declared in the system if the Q statistic value is higher than threshold value (Q_{α}) for new data set.

The following section demonstrates the implementation of the fault detection methods described above and analyzes the effectiveness of all techniques.

4. Illustrative examples

The effectiveness of the kernel extensions of PCA and PLS for fault detection purposes will be demonstrated through two illustrative nonlinear examples, using a simulated synthetic data set and a simulated continuous stirred tank reactor (CSTR).

4.1. Simulated synthetic data

Synthetic nonlinear data can be simulated through the following model [37]:

$$\begin{aligned} x_1 &= u^2 + 0.3\sin(2\pi u) + \varepsilon_1 \\ x_2 &= u + \varepsilon_2 \\ x_3 &= u^3 + u + 1 + \varepsilon_3 \end{aligned} \tag{61}$$

where u is a variable that is defined between -1 and 1 and ε_i is a variable of independent white noise distributed uniformly between -0.1 and 0.1. Training and testing data sets of 401 observations each are generated using the model above. The performance of KPCA and KPLS techniques is illustrated and compared to the conventional PCA and PLS methods for two different cases. In the first case, the sensor measuring the first variable x_1 is assumed to be faulty with a single fault. In the second case, multiple faults are assumed to occur simultaneously in x_1, x_2, and x_3.

Figure 3 shows the generated data.

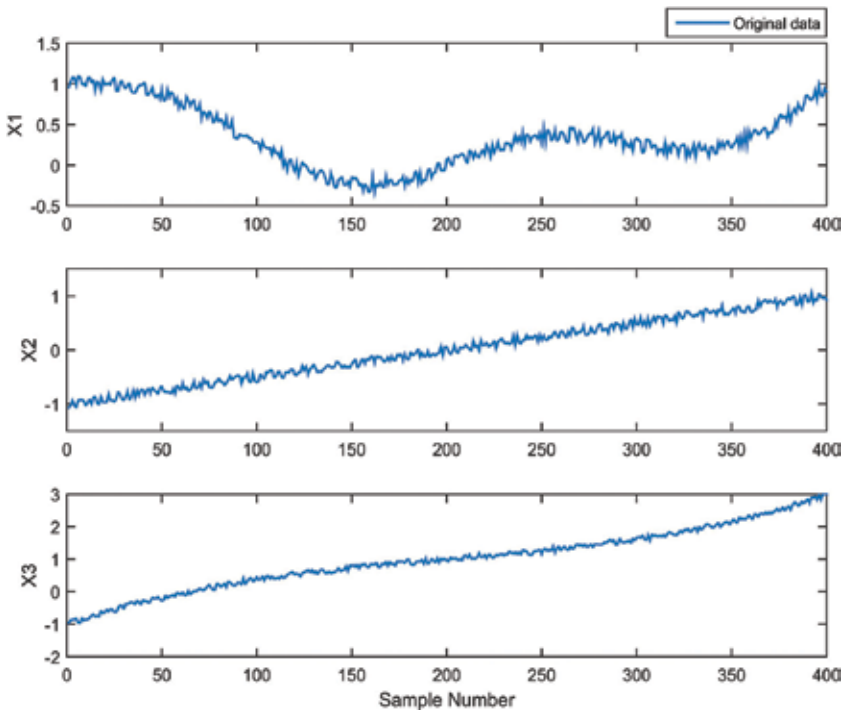


Figure 3. Generated data.

Case 1

In this case, a single fault of magnitude unity is introduced between observations 200 and 250 in x_1 in the testing data set. The Gaussian kernel was chosen to model the nonlinearity in the process data. The most common fault detection metrics used are the missed detection rate, the false alarm rate, and the out-of-control average run length (ARL_1). The missed detection rate is

when a fault goes undetected in the faulty region, while the false alarm rate is when an observation is flagged as a fault in the non-faulty region. The false alarm and missed detection rates are also commonly referred to as Type I and Type II errors, respectively. ARL_1 is the number of observations, and it takes for a particular technique to flag a fault in faulty region and is used to assess the speed of a detection. The fault-free and faulty data are shown in **Figures 4** and **5**, respectively.

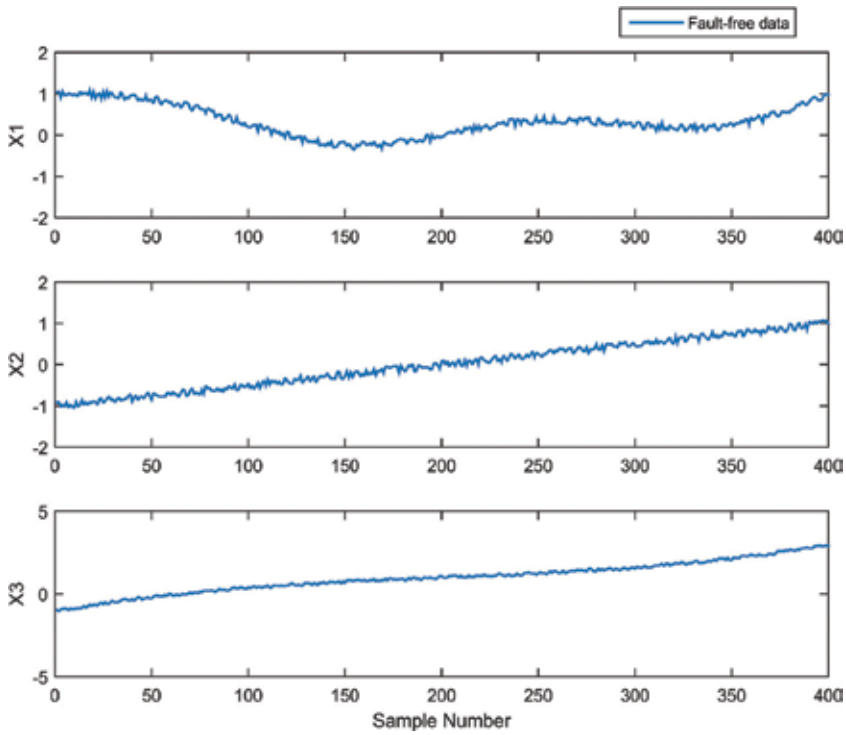


Figure 4. Fault-free data.

The fault detection (FD) performance of PCA-, KPCA-, PLS-, and KPLS-based Q methods is shown in **Figures 6** and **7** as well as **Table 1**. The results show that both KPCA and KPLS-based Q provide a better FD performance than the linear PCA- and PLS-based Q methods and are able to detect the faults with lower missed detection rates, false alarm rates, and ARL_1 values (see **Table 1**).

Case 2

In this case, a multiple faults of magnitude unity are introduced between observations 200 and 250 in x_1, 100 and 150 in x_2, and 385 and 401 in x_3 in the testing data set (as shown in **Figure 8**).

The FD performance of the kernel PCA and kernel PLS methods is illustrated and compared to that of the conventional PCA and PLS methods using the Q statistic. The Q statistic was chosen for analysis, since it is often better able to detect smaller faults using the residual space and for simplicity of analysis as well. The fault detection performance of a particular process monitoring technique can be monitored using multiple fault detection metrics.

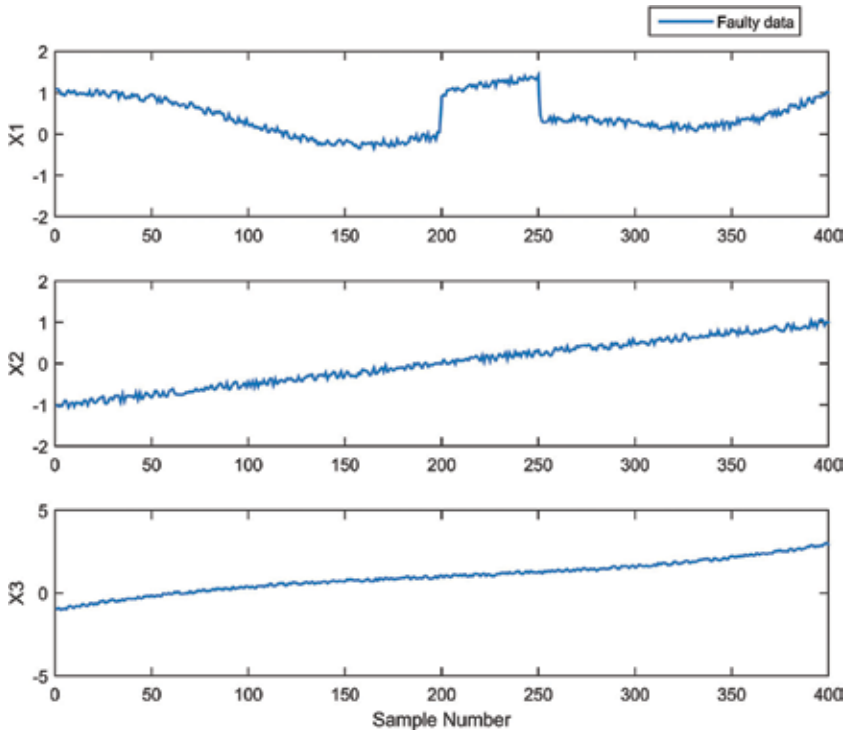


Figure 5. Faulty data in the presence of single fault in x_1.

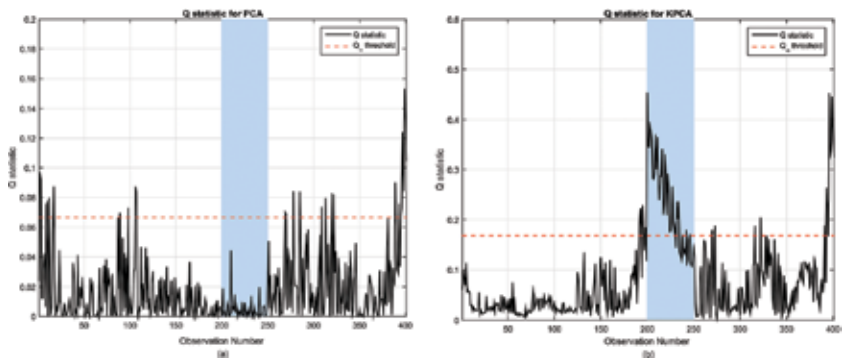


Figure 6. Monitoring single fault using PCA- and KPCA-based Q methods—case 1.

As can be seen through **Figures 9(a)** and **10(a)** and **Table 2**, the conventional linear PCA and PLS techniques are unable to effectively capture the nonlinearity present in the data set, which leads to entire sets of faults going undetected for both the linear PCA and PLS techniques. However, as demonstrated in **Figures 9(b)** and **10(b)**, the KPCA and KPLS-based Q techniques are better able to detect the faults with lower missed detection rates, false alarm rates, and ARL_1 values than the linear PCA and PLS methods (as shown in **Table 2**). These improved results can be attributed to the fact that the kernel techniques are able to capture the nonlinearity

http://dx.doi.org/10.5772/67347

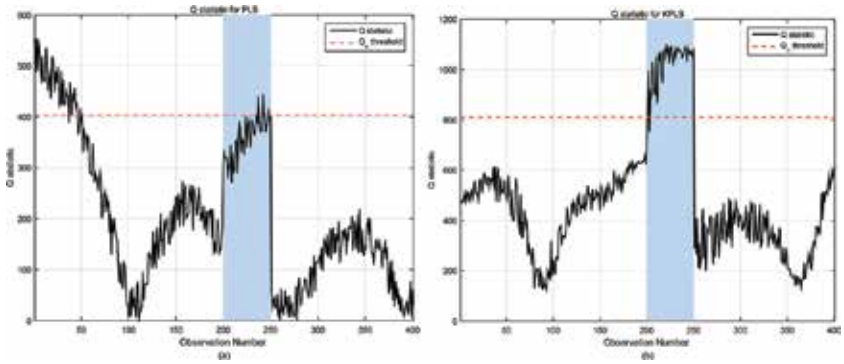


Figure 7. Monitoring single fault using PLS- and KPLS-based Q methods—case 1.

	Missed detection (%)	False alarm (%)	ARL_1
PLS-based Q statistic	90.1961	13.1429	36
KPLS-based Q statistic	3.9216	0	2
PCA-based Q statistic	100	7.4286	-
KPCA-based Q statistic	27.4510	5.1429	1

Table 1. Summary of missed detection (%), false alarms (%), and ARL1 for case 1.

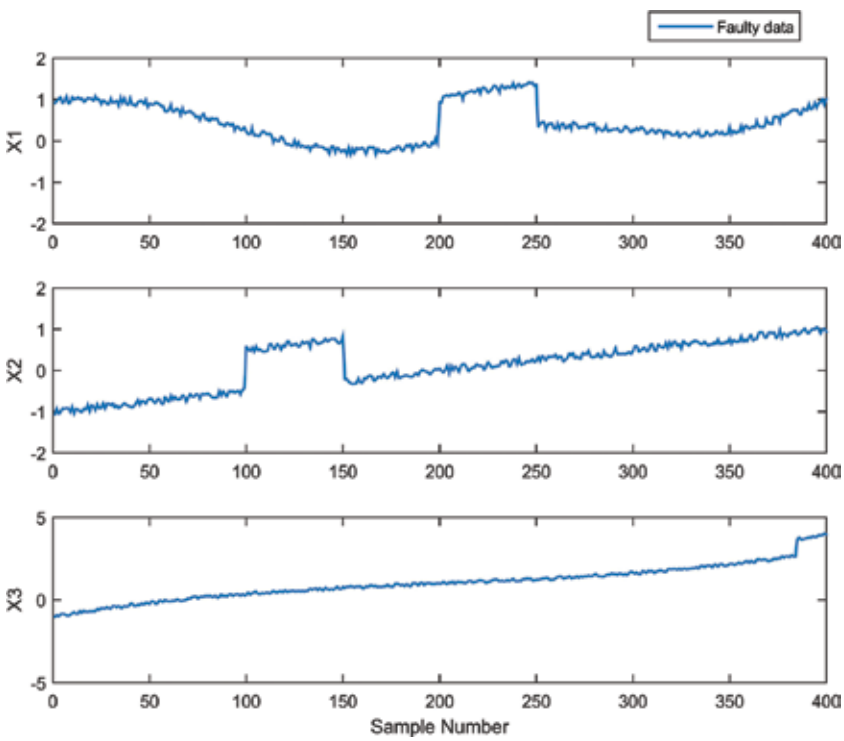


Figure 8. Faulty data in the presence of multiple faults in x_1, x_2, and x_3.

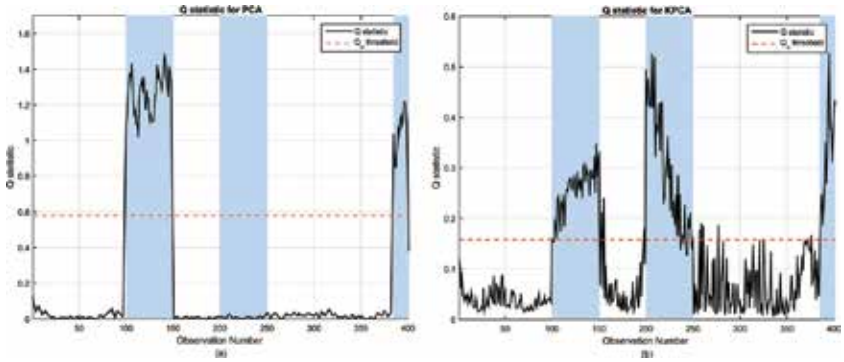


Figure 9. Monitoring multiple faults using PCA- and KPCA-based Q methods using simulated synthetic data—case 2.

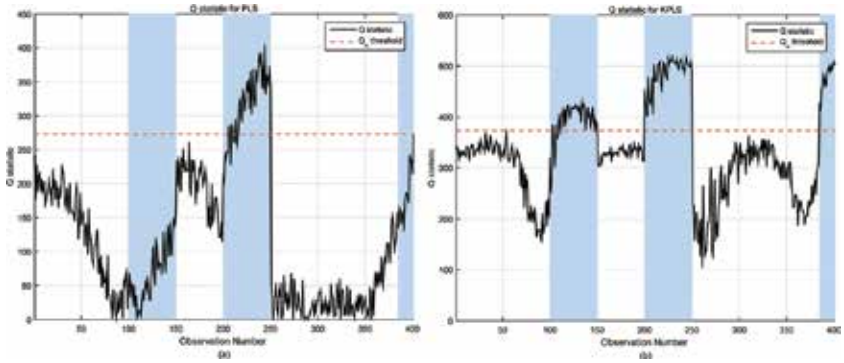


Figure 10. Monitoring multiple faults using PLS- and KPLS-based Q methods using simulated synthetic data—case 2.

	Missed detection (%)	False alarm (%)	ARL_1
PLS-based Q statistic	67.2269	0	58
KPLS-based Q statistic	6.7227	0	1
PCA-based Q statistic	42.8571	0	1
KPCA-based Q statistic	8.4034	2.8369	1

Table 2. Summary of missed detection (%), false alarms (%), and ARL1 for case 2.

in the hyperdimensional feature space, providing better detection especially in this case where there are multiple faults in the system.

4.2. Simulated CSTR model

In order to effectively assess the performance of the kernel PCA and kernel PLS techniques, it is also necessary to examine the performance of the techniques using an actual process application as well. A continuous stirred tank reactor model can be used to generate nonlinear data, and the fault detection charts can be applied to test their performance.

http://dx.doi.org/10.5772/67347

4.2.1. CSTR process description

The dynamic for the CSTR that was utilized for this simulated example is represented as follows [5]:

$$\begin{aligned}\frac{\partial C_A}{\partial t} &= \frac{F}{V}(C_{A_0} - C_A) - k_0 e^{-E/RT} C_A \\ \frac{\partial T}{\partial t} &= \frac{F}{V}(T_0 - T) + \frac{(-\Delta H)}{\rho C_P} e^{-E/RT} C_A - \frac{q}{V \rho C_p} \\ q &= \frac{a F_c^{b+1}}{F_c + \left(\frac{a F_c^b}{2 \rho_c C_{pc}}\right)} (T - T_{cin})\end{aligned} \tag{62}$$

where k_0, E, F, and V represent the reaction rate constant, activation energy, flow rates (both inlet and outlet), and reactor volume, respectively. The concentration of A in the inlet stream and of B in the exit stream is represented by C_A and C_B, respectively. The temperatures of the inlet stream and of the cooling fluid in the jacket are T_i and T_j, respectively. ΔH, U, A, ρ, and C_p represent the heat of reaction, overall heat transfer coefficient, area through which the heat transfers to the cooling jacket, density, and heat transfer coefficient of all streams, respectively.

Using the described CSTR model, 1000 observations were generated, which was assumed to be initially noise-free. Zero-mean Gaussian noise with a signal-to-noise ratio of 20 was used to contaminate the noise-free process observations, in order to replicate reality. **Figure 11** shows

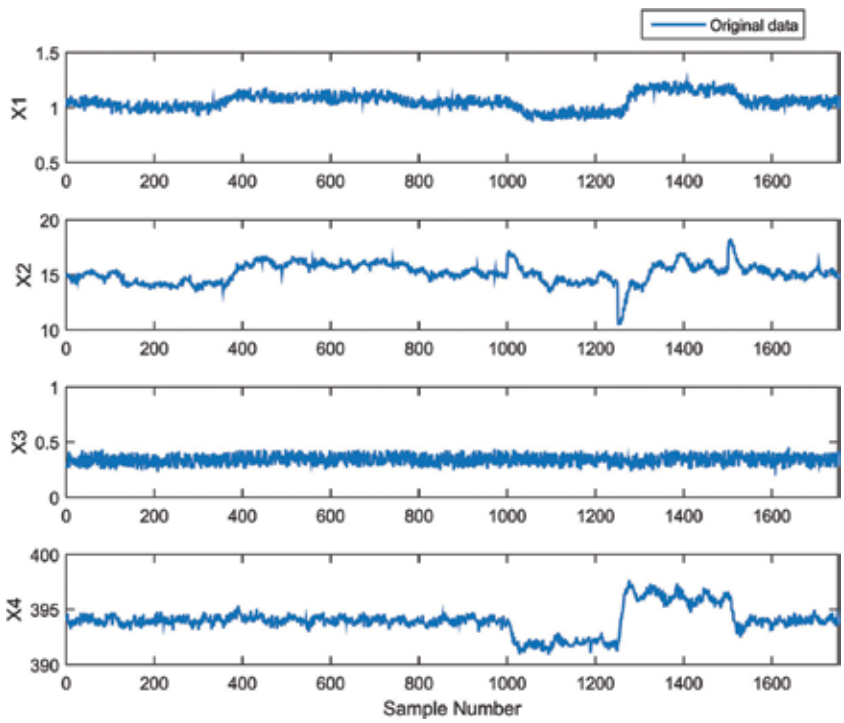


Figure 11. Generated continuously stirred tank reactor (CSTR) data.

the generated CSTR data. This data set was then split into training and testing data sets, of 500 observations each. Faults of magnitude 3σ were added to the temperature and concentration process variables in the testing data set, at three different locations: observations 101–150, 251–350, and 401–450. σ is the standard deviation of that particular process variables. **Figures 12** and **13** show the unfaulty and faulty data, respectively. Similar to the previous example, the performance of kernel PCA and kernel PLS methods is compared to the conventional linear PCA and PLS methods using the Q statistic.

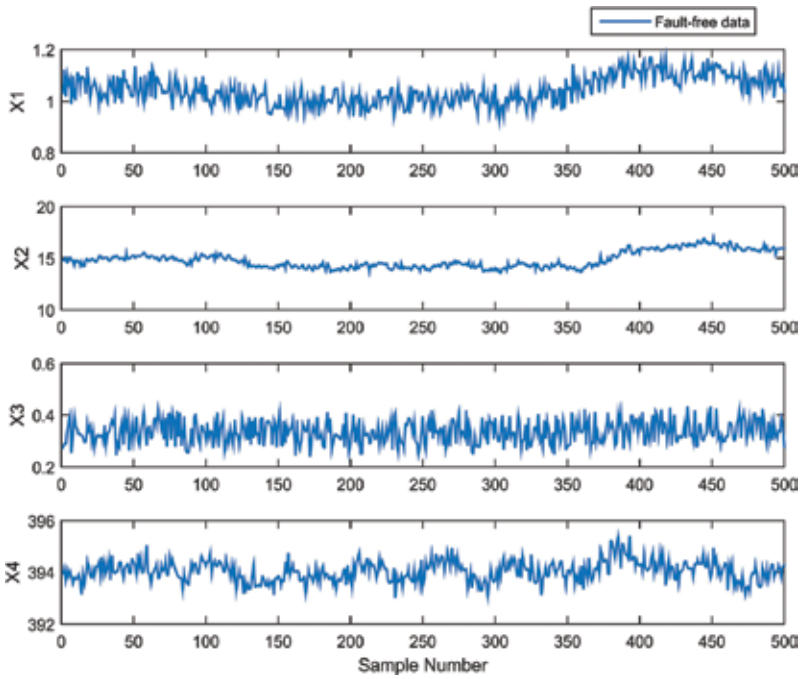


Figure 12. Fault-free data.

For this example, comparing the two conventional techniques, we can see that the PCA-based Q statistic is unable to all faults (see **Figure 14** (**a**)), while the PLS-based Q model is able to better detect the faults (see **Figure 15** (**a**)). However, the kernel PCA and kernel PLS-based Q techniques are able to provide result charts with lower missed detection rates, false alarm rates, and ARL_1 values than their corresponding conventional techniques (see **Figures 14(b)** and **15(b)**). These improved results can once again be attributed to the kernel techniques being able to effectively capture the nonlinearity of the data in the hyperdimensional feature space. The FD results using the two examples showed that the kernel PLS-based Q provides a relative performance compared to the kernel PCA Q. This is because kernel PCA is an input space model and cannot take into consideration outcome measures and most chemical processes or many of them are usually described by input-output space models.

http://dx.doi.org/10.5772/67347

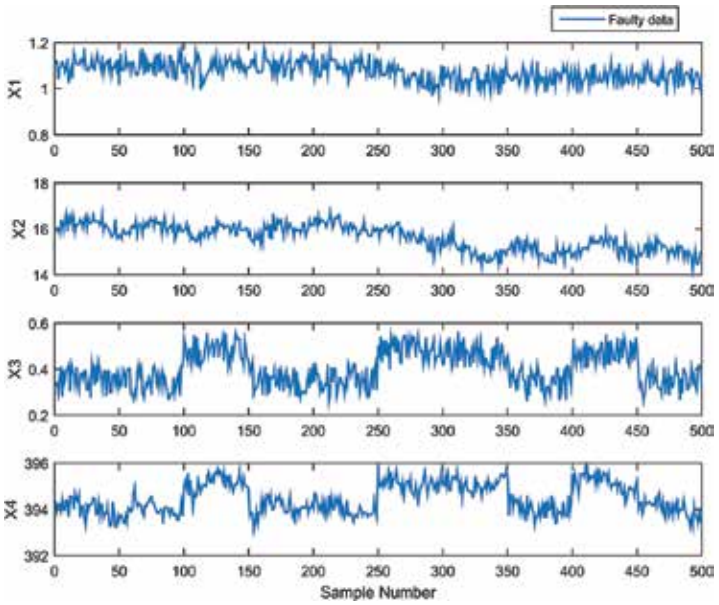


Figure 13. Faulty data in the presence of multiple faults in temperature and concentration.

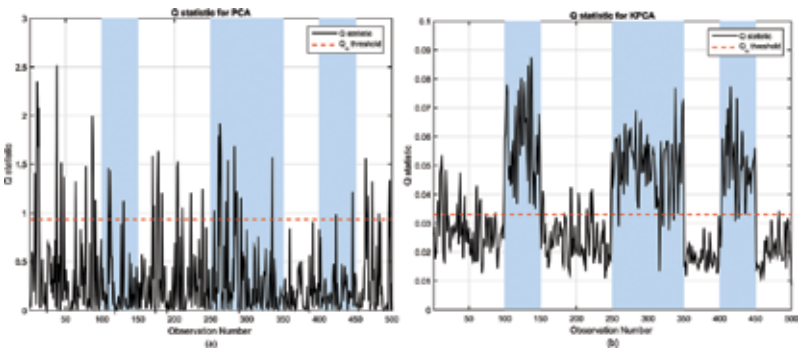


Figure 14. Fault detection using PCA- and kernel PCA-based Q methods using CSTR data.

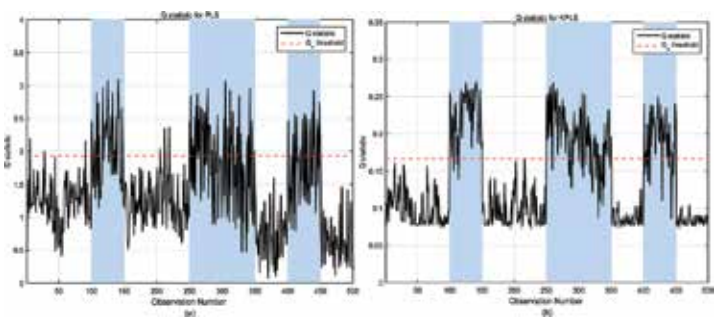


Figure 15. Fault detection using PLS- and kernel PLS-based Q methods using CSTR data.

5. Conclusion

In this chapter, a nonlinear multivariate statistical techniques are used for fault detection. Kernel PCA and kernel PLS have been widely used to monitor various nonlinear processes, such as distillation columns and reactors. Thus, in the current work, both kernel PCA and kernel PLS methods are used for nonlinear fault detection of chemical process. A commonly used fault detection index is Q-square statistic, and it is used to detect fault in the system. The fault detection performance using linear and nonlinear input models (PCA and kernel PCA) and input-output models (PLS and kernel PLS) is evaluated through two simulated examples, synthetic data set and continuous stirred tank reactor (CSTR). Missed detection rate, false alarm rate, and ARL_1 are the parameters used to compare the fault detection techniques. The results of the two case studies showed that the kernel PCA and kernel PLS-based Q provide improved fault detection performance compared to the conventional PCA- and PLS-based Q methods.

6. Acknowledgements

This work was made possible by NPRP grant NPRP7-1172-2-439 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

Author details

Mohammed Ziyan Sheriff[1,2], Chiranjivi Botre[1], Majdi Mansouri[3], Hazem Nounou[3], Mohamed Nounou[2*] and Mohammad Nazmul Karim[1]

*Address all correspondence to: mohamed.nounou@qatar.tamu.edu

1 Artie McFerrin Department of Chemical Engineering, College Station, TX, USA

2 Chemical Engineering Program, Texas A&M University at Qatar, Doha, Qatar

3 Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar

References

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis part I: Quantitative model-based methods," *Comput. Chem. Eng*., vol. 27, pp. 293–311, 2003.

[2] V. Venkatasubramanian, R. Rengaswamy, and S. N. Ka, "A review of process fault detection and diagnosis part II: Qualitative models and search strategies," *Comput. Chem. Eng*., vol. 27, pp. 313–326, 2003.

[3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part III: Process history based methods," *Comput. Chem. Eng*., vol. 27, pp. 327–346, 2003.

[4] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res*., vol. 52, no. 10, pp. 3543–3562, 2013.

[5] M. Mansouri, M. Nounou, H. Nounou, and N. Karim, "Kernel PCA-based GLRT for nonlinear fault detection of chemical processes," *J. Loss Prev. Process Ind*., vol. 40, pp. 334–347, 2016.

[6] Jolliffe, Ian. *Principal component analysis*. John Wiley & Sons, Ltd., 2002.

[7] M. F. Harkat, G. Mourot, and J. Ragot, "An improved PCA scheme for sensor FDI: Application to an air quality monitoring network," *J. Process Control*, vol. 16, no. 6, pp. 625–634, 2006.

[8] J. P. George, Z. Chen, and P. Shaw, "Fault detection of drinking water treatment process using PCA and Hotelling's T 2 chart," pp. 970–975, 2009.

[9] J. Yu, "Fault detection using principal components-based gaussian mixture model for semiconductor," *IEEE Trans. Semicond. Manuf*., vol. 24, no. 3, pp. 432–444, 2011.

[10] Y. Zhang, W. Du, Y. Fan, and L. Zhang, "Process fault detection using directional kernel partial least squares," *Ind. Eng. Chem. Res*., vol. 54, no. 9, pp. 2509–2518, 2015.

[11] P. Nomikos and J. F. MacGregor, "Multi-way partial least squares in monitoring batch processes," *Chemom. Intell. Lab. Syst*., vol. 30, no. 1, pp. 97–108, 1995.

[12] T. Kourti and J. F. J. F. MacGregor, "Process analysis, monitoring and diagnosis, using multivariate projection methods," *Chemom. Intell. Lab. Syst*., vol. 28, pp. 3–21, 1995.

[13] M. Mansouri, M. Z. Sheriff, R. Baklouti, M. Nounou, H. Nounou, A. Ben Hamida, and M. N. Karim, "Statistical fault detection of chemical process - comparative studies," *J. Chem. Eng. Process Technol*., vol. 7, no. 1, pp. 1–10, 2016.

[14] S. Joe Qin, "Statistical process monitoring: basics and beyond," *J. Chemom*., vol. 17, no. 8–9, pp. 480–502, 2003.

[15] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associted with principal component analysis," *Technometrics*, vol. 21, no. 3, pp. 341–349, 1979.

[16] J. E. Jackson "Quality control methods for several related variables," *Technometrics*, vol. 1, no. 4, pp. 359–377, 1959.

[17] M. Zhu and A. Ghodsi, "Automatic dimensionality selection from the scree plot via the use of profile likelihood," *Comput. Stat. Data Anal*., vol. 51, no. 2, pp. 918–930, 2006.

[18] G. Diana and C. Tommasi, "Cross-validation methods in principal component analysis: A comparison," *Stat. Methods Appl.*, vol. 11, no. 1, pp. 71–82, 2002.

[19] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space," *J. Mach. Learn. Res.*, vol. 2, pp. 97–123, 2002.

[20] P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Anal. Chim. Acta*, vol. 185, pp. 1–17, 1986.

[21] B. S. Dayal and J. F. MacGregor, "Recursive exponentially weighted PLS and its applications to adaptive control and prediction," *J. Process Control*, vol. 7, no. 3, pp. 169–179, 1997.

[22] S. J. Qin, "Recursive PLS algorithms for adaptive data modeling," *Comput. Chem. Eng.*, vol. 22, no. 4–5, pp. 503–514, 1998.

[23] J. F. MacGregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process monitoring and diagnosis by multiblock PLS methods," *AIChE J.*, vol. 40, no. 5, pp. 826–838, 1994.

[24] T. Kourti, P. Nomikos, and J. F. MacGregor, "Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS," *J. Process Control*, vol. 5, no. 4, pp. 277–284, 1995.

[25] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.

[26] U. Krüger and L. Xie, *Statistical monitoring of complex multivariate processes: With applications in industrial process control*. Chichester, West Sussex; Hoboken, N.J: Wiley, 2012.

[27] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.

[28] S. Wold, N. Kettaneh-Wold, and B. Skagerberg, "Nonlinear PLS modeling," *Chemom. Intell. Lab. Syst.*, vol. 7, no. 1–2, pp. 53–65, 1989.

[29] Rosipal, R. (2011). Nonlinear partial least squares: An overview. In Lodhi H. and Yamanishi Y. (Eds.), *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, pp. 169–189, 2011. ACCM, IGI Global. Retrieved from http://aiolos.um.savba.sk/~roman/Papers/npls_book11.pdf

[30] 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," *Chemom. Intell. Lab. Syst.*, vol. 75, no. 1, pp. 55–67, 2005.

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

[32] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

http://dx.doi.org/10.5772/67347

[33] J. L. Godoy, D. A. Zumoffen, J. R. Vega, and J. L. Marchetti, "New contributions to nonlinear process monitoring through kernel partial least squares," *Chemom. Intell. Lab. Syst*., vol. 135, pp. 76–89, 2014.

[34] R. Rosipal, "Kernel partial least squares for nonlinear regression and discrimination," *Neural Netw. World*, vol. 13, no. 3, pp. 291–300, 2003.

[35] M.-F. Harkat, S. Djelel, N. Doghmane, and M. Benouaret, "Sensor fault detection, isolation and reconstruction using nonlinear principal component analysis," *Int. J. Autom. Comput*., vol. 4, no. 2, pp. 149–155, 2007.

[36] S. A. Vejtasa and R. A. Schmitz, "An experimental study of steady state multiplicity and stability in an adiabatic stirred reactor," *AIChE J*., vol. 16, no. 3, pp. 410–419, 1970.

[37] Botre, Chiranjivi, et al. "Kernel PLS-based GLRT method for fault detection of chemical processes," *J. Loss Prevent. Process Industries*, 43 (2016): 212–224.

Chapter 11

Noninvasive Methods for Condition Monitoring and Electrical Fault Diagnosis of Induction Motors

Muhammad Aman Sheikh, Nursyarizal Mohd Nor,

Taib Ibrahim, Sheikh Tahir Bakhsh,

Muhammad Irfan and Hanita Binti Daud

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67245

Abstract

This chapter provides a comprehensive analysis of noninvasive methods to diagnose stator winding insulation faults of an induction motor. Further, a novel noninvasive method is proposed to diagnose the root cause of winding failure due to unbalanced voltage to avoid catastrophic failure. Therefore, a winding function approach is utilized to derive an analytical expression for stator winding distribution and magnetomotive force (MMF). This tactic qualifies the conductor segment that generates MMF, and it also helps to analyze a healthy current spectrum. One can easily observe higher order harmonics in current spectrum; therefore, a new series of rotor harmonics is introduced to diagnose unbalanced supply. The locus of these harmonics is dependent on the poles, rotor bars, and slip. Due to the rapid complexity in industrial plants, it is inconceivable to continue human inspection to diagnose the faults. Thus, to avoid human inspection, in addition to new series of rotor harmonic, a fully automatic method based on neural network is proposed. This method not only diagnoses unbalanced voltage but it also recognize the percentage of unbalanced voltage by use of feed-forward multilayer perceptron (MLP) trained by back propagation. Finally, the experimental results shows the validation of this research work proposed method.

Keywords: artificial neural network (ANN), diagnosis, fault detection, induction machine, interturn short circuit, rotor harmonic, unbalanced voltage

1. Introduction

Condition monitoring is of a great concern to ensure continue production, reliability, and consequently, avoid catastrophic failures. Therefore, online condition monitoring of induction motor has become a subject of interest and a challenging task for the protection of a motor [1, 2]. In induction motor, the most critical component, which is also considered the main source of motor failures, is stator winding insulation failure [3–6]. Over the last few decades, numerous surveys have been carried out on the reliability of induction motor. It has been experienced that 37% of motor failure is due to insulation breakdown.

It is found that unbalanced voltage is a major source of insulation breakdown. The unbalanced voltage results in an unscheduled shutdown of the process and causes enormous costs. Thus, it is desirable that the misalignment in the voltage source, which can result in a severe failure, should be identified at an incipient stage. A substantial amount of work has been proposed based on current and voltage [7], current residue, the phase shift in current and voltage [8], and sequence components [4, 9–11]. The majority of the methods to diagnose the unbalanced voltage essentially (invasive or noninvasive) monitor single or multiple indicators of stator faults. Most of the latter are extracted from a signature analysis. Therefore, it is important to be familiar with the spectrum of a healthy motor, which is a challenging task due to the complexity of electromagnetic processes, time, and space variables of an induction motor. Conversely, it is essential to have a mathematical modeling that enables to extract a geometrical representation of the motor. Beside this, to interpret the results from the mathematical model it is necessary to introduce an appropriate analytical model. The modeling of stator conductor segments and MMF is a most reliable approach to execute correct interpretation of induction motor.

As mentioned earlier that the unbalanced voltage is a major source of insulation breakdown, therefore, the aim of this paper is to diagnose this fault at an incipient stage. The unbalanced voltage can lead to turn-turn fault, phase-phase fault, phase-ground fault, and even a destructive effect on the stator core. During the last few decades, the condition monitoring of induction motor has become a vigorous area of research. Substantial work has been executed to propose and develop various techniques and methods to diagnosis motor faults. Through most relevant literature, three major classes of methods are discovered which support the task of fault monitoring [8]. The first class is based on a signal analysis [12, 13] that uses spectra in a frequency domain, time domain, time-frequency domain, and high-order harmonics. The second approach is based on analytical modeling [14] that involves mathematical models to measure input and output feature such as residuals, state estimation, and parameter estimation that incorporate the artificial intelligence (AI) to online automate analyze the health of induction motor through measured signals [15]. In engineering the AI tools are of great significance to solve various complex problems [16, 17]. The AIs are classified into four different groups as an expert system, genetic algorithm (GA), fuzzy logic, and artificial neural networks (ANNs) [18].

This paper aims to analyze the current spectrum of the induction motor to diagnose and distinguish between balanced and unbalanced voltage against the different level of unbalanced voltage. In the first part of the chapter, a comprehensive analysis of noninvasive methods is

http://dx.doi.org/10.5772/67245

presented to diagnose electrical faults. While in the second part, an analytical model is derived for the conductor segments, which are responsible for the generation of MMF. In third part, a reminiscence is carried out on MMF and an expression is derived to modify MMF with reference to the slot conductors that play an import role in the generation of MMF by introducing all harmonics terms that have influence in MMF waves. Further, a novel noninvasive technique based on ANN is presented in Section 6 along with results and test procedure to diagnose and distinguish unbalanced voltage at incipient stage in order to avoid catastrophic failure. Finally, conclusions and future work are presented in Section 7.

2. Noninvasive stator fault diagnosis methods

Diagnosis of stator winding insulation failure has been extensively studied. Some of the approaches are based on laboratory tests while other are based on experimental results and analysis through simulation. Mostly, the researchers divide the diagnosis techniques into two main categories, i.e., invasive and noninvasive. Although invasive fault diagnosis methods help to diagnose the defect but it has many drawbacks, some are mentioned as: most of the invasive techniques are offline or sensor bases, for installation of the sensors the normal operation of an induction motor is disturbed. Moreover, by influenced operation of contiguous equipment, there is no proper capability to diagnosis the weak faults. In contrast, noninvasive methods make use of stator current and voltage, or other important electrical quantities. There is no constraint of sensors to be installed. It is a cost-effective approach because in industries, current and voltages are usually measured, thus the economic cost of current and voltage measurement is not imposed on the system. Therefore, noninvasive methods in recent years are preferred over invasive. The next subsection describes the operating principle of noninvasive indexes of how the stator winding faults are diagnosed.

2.1. The air-gap torque

The air-gap torque between the stator winding and the rotor is induced by current and flux linkages that are sensitive to asymmetry caused by unbalanced voltage supply. The fault can be diagnosed by stator current and voltage. The motor with interturn short circuit will generate forward and backward rotational fields; the forward rotational field has a constant torque with correspondence to the motor rotating field, while the backward rotational field has twice the oscillating torque as compared to motor supplied frequency [19]. Therefore, the frequency of backward rotating field may be used as an interturn fault index. However, the induction motor with unbalanced three-phase voltages also accesses the same oscillating torque component. This index may be an expensive approach because it requires measuring both current and voltage. Moreover, this index cannot distinguish between interturn and unbalanced voltage.

2.2. The stator line current

The stator line current is the index that may be used to reflect a variety of noninvasive indexes to diagnose interturn fault. The stator line current indexes are briefly described in the succeeding

text. First, consider three current sequences positive, negative, and zero of a three phase as follows:

$$X^0 = \frac{1}{3}(X_a + X_b + X_c) \tag{1}$$

$$X^+ = \frac{1}{3}(X_a + aX_b + a^2X_c) \tag{2}$$

$$X^- = \frac{1}{3}(X_a + a^2X_b + aX_c) \tag{3}$$

where $a = e^{j2\pi/3}$. The healthy induction motor without any fault or symmetry has zero current for a negative sequence, whereas the interturn fault in motor generates considerable negative sequence current. While there is minute change observed for positive sequence current, there is no effect of load variation on the negative sequence current, but it is highly sensitive to the imbalance of the supply voltages and can easily inherent asymmetry [19]. Therefore, this index does not properly indicate the interturn fault and unbalanced voltage.

Park vector estimation, analysis uses the machine line current to diagnose interturn faults [20]. Under healthy condition, the motor Park vector pattern corresponds to a circular pattern. While in the occurrence of interturn short circuits, the Park vector pattern manifests to an elliptic representation and with the severity of fault the ellipticity increases. Through experiments and simulation results it has been proven that Park vector is a reliable approach to detect the interturn fault but does not investigate unbalanced voltage supply [20, 21].

In induction motor, the motor current signature analysis (MCSA) is the most extensively used diagnosis index [22, 23]. It is observed that current spectrum is also affected by unbalanced voltage supply. Thus, the current spectrum will have third-order harmonics under interturn as well as under unbalanced conditions. Therefore, MSCA is confused with the similar behavior of induction motor harmonics under faulty and unbalanced conditions. Hence, MCSA approach is not a reliable index to sufficiently identify the faults.

2.3. Power

The instantaneous electrical power is one of the reliable faults diagnose indexes for induction motor [24], and some of the power frequency components have been identified as the stator windings fault index [8]. Power frequencies are only dependent on the severity of the fault but independent of the speed. The major factor that makes this index less reliable is that it is influenced by peripheral factors such as inherent asymmetries, unbalanced supply voltage, measuring equipment, and load variations. It is an expensive fault index because this requires measurement of both voltage and current signals.

2.4. Sequence impedance matrix

Negative sequence apparent impedance is a noninvasive method and is used as a fault index to detect stator winding faults [25, 26]. The impedance is obtained through the phasor voltage

and phasor current. When interturn fault occurs, the impendence of the motor will be disturbed. In this way, the change in the impedance is an indication of the faulty motor. This index is less sensitive to load variation and unbalanced voltage. Therefore, it can be a preferable method to diagnose the interturn stator winding faults, but it is a bit expensive and lengthy approach because both the signal voltage and current have to be monitored.

2.5. Stator voltage

Stator voltage is also used as an index to detect the stator winding faults [27, 28]. By adopting this approach, the algebraic sum of three-phase voltages is useful because this approach relies on the variation of residual voltage caused by an interturn fault. For healthy motor, the residual voltage is zero, while it is nonzero for an interturn motor. In a real scenario even in the healthy induction motor, there are no balanced three-phase windings. Therefore, this index is less reliable under load conditions. In this method, machine voltage harmonics are observed by disconnecting the terminals [27], so this method cannot be applied for online condition monitoring. After disconnecting the voltage terminals, the odd harmonics of the motor are observed and used as interturn fault index [28].

From the above-mentioned methods, none of the methods is capable of distinguishing unbalanced voltage. Thus, to tackle unbalanced anomalies at the incipient stage all the three classes of conditioning monitoring techniques are implemented in this research to diagnose unbalanced voltage at incipient stage, i.e., signal analysis class, analytical modeling class, and knowledge-based class.

3. Analytical expression for spatially distributed stator winding

In induction motors, the armature winding is at the stator side between the stamped slots. **Figure 1** shows a stator with 12 slots and it can be illustrated that the winding in an induction motor is not a single coil but is composed of a number of coils that are spatially distributed and placed in the axial direction to provide room for the generation of magnetic flux.

Moreover, the stator winding of an induction motor is distributive in nature and can be elucidated either by a discrete or by a continuous formulation. The discrete description refers to the number of conductors in a slot while continue distribution is based on the density of the windings. However, the conductors of stator winding are not placed in a continuous manner, but rather are located in the slots. As the conductors have a physical size, which could be viewed with sufficient resolution, therefore, the stator winding can be described mathematically to derive an analytical expression for the conductor segments that are responsible for the generation of MMF that is the first task of proposed work and each slot is partition down to slot and end conductors as shown in **Figure 2**.

The main focus of this section is to derive an analytical model for the slot conductors that are oriented in axial direction. The reason to focus more on slot conductors is that these windings are the portion of stator conductor that establishes MMF and that is involved in the production of a torque. **Figure 1(a)** illustrates the stator of induction motor with 12 slots, whereas the

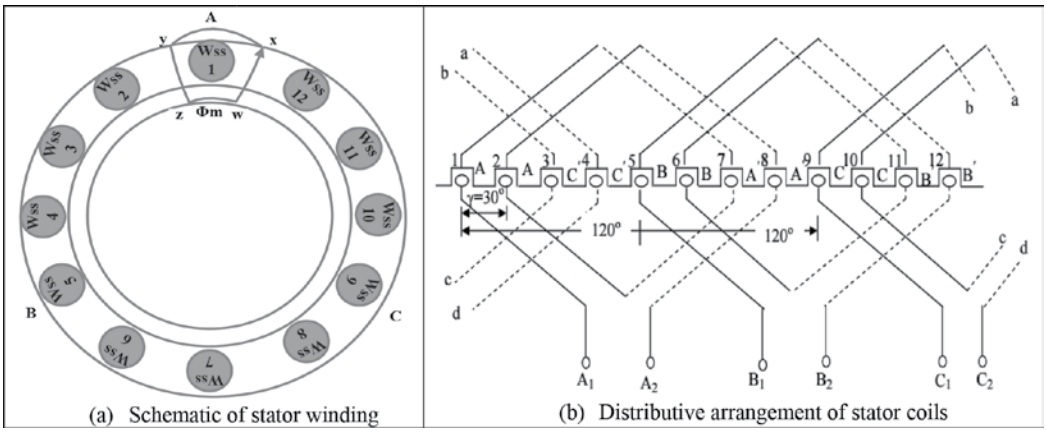


Figure 1. Stator winding arrangement in the slots.

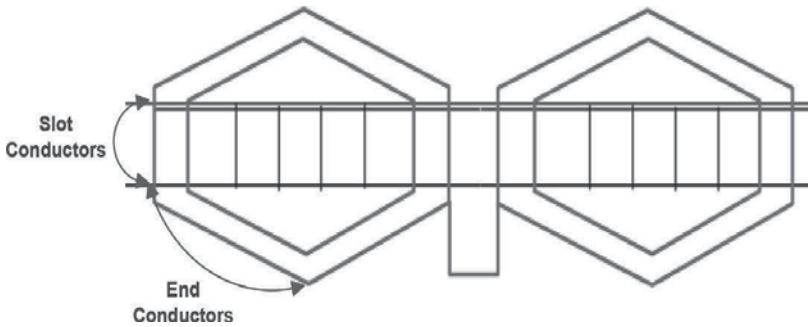


Figure 2. Partition of stator winding into slot and end conductions.

stator windings are located in each slot. The notation $W_{ss,i}$ indicates the conductors in the ith slot where "*as*" refer to phase-a of the stator winding. The conductors in the slots are shown as open circles. Subsequently, the notation can be used to derive an expression for the segment of slot conductors. For more generalize case, generalizing this notation, $W_{ss,i}$ is the conductors of any phase in slot ith of winding:

$$W_{ss} = \sum_{i=1}^{n} W_{ss,i} u(W_{ss,i}) \tag{4}$$

Figure 3 is a version of a developed diagram of the stator winding of induction motor. In this developed diagram, instead of front or back view of a machine, the angle of vision is from the center of a stator to outward. Within the slots, $W_{x,i}$ denotes the number of slot conductors of i^{th} *slot*. Where, W_{os} and W_{is} are the variables representing front end conductor directed to the left or right, respectively. It can be illustrated that for a symmetrical induction motor without any fault, these conductors are same in numbers as given in expression:

http://dx.doi.org/10.5772/67245

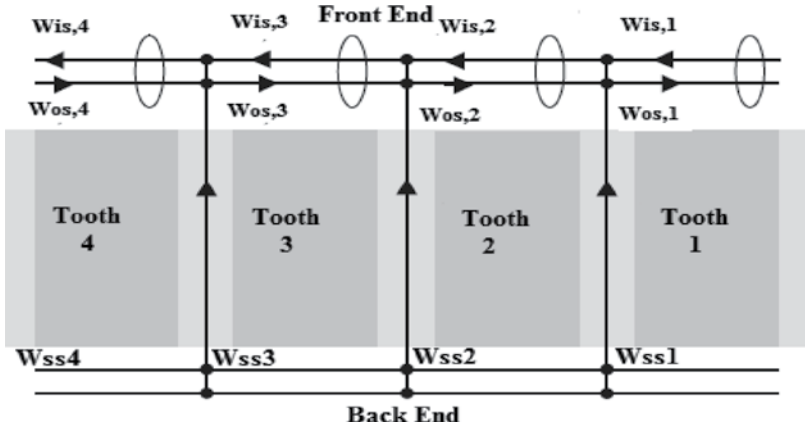


Figure 3. A version center view of stator through a developed diagram.

$$\sum_{i=1}^{n} Wo_{ss,i} u(W_{ss,i}) = \sum_{i=1}^{n} Wi_{ss,i} u(W_{ss,i}) \tag{5}$$

In most of the induction motors, the stator winding is multilayered that contains conductors from multiple windings (phases) as shown in **Figure 4**. Conductors out of the slot are known as end conductors and are of importance because they contribute to the inductance and resistance of the winding. Thus, total end conductors are given as in Eqs. (6)–(7).

$$E_{ss} = \sum_{i=1}^{n} Wo_{ss,i} + \sum_{i=1}^{n} Wi_{ss,i} \tag{6}$$

$$E_{ss} = \sum_{i=1}^{n} E_{ss,i} u(E_{ss,i}) \tag{7}$$

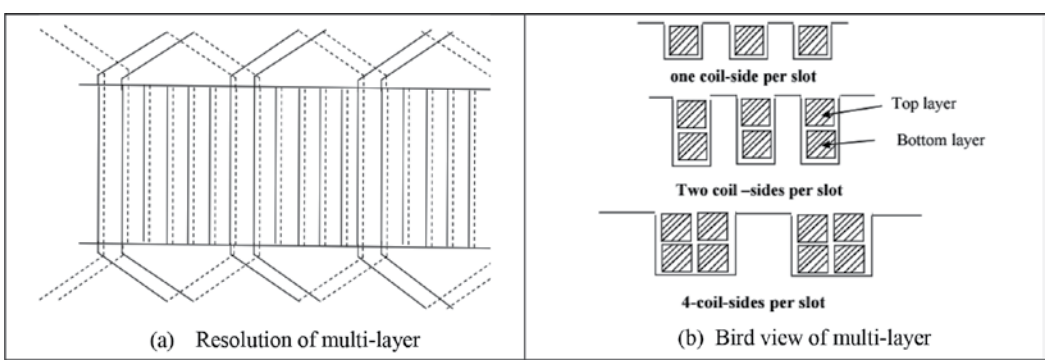


Figure 4. Multilayer stator winding. (a) Resolution of multilayer. (b) Bird view of multilayer.

As the stator winding is segmented into slot and end conductors, therefore the total number of turns associated with the winding will be the combination of slot conductors and end conductors:

$$W_T = \sum_{i=1}^{n} W_{ss,i} u(W_{ss,i}) + \sum_{i=1}^{n} E_{ss,i} u(E_{ss,i}) \tag{8}$$

Further, to calculate a center of stator teeth and stator slot, respectively, the following expression can be used based on the developed diagram:

$$\varnothing_{ss,i} = \frac{\pi(2i-2)}{T_s} + \varnothing_{ss,1} \tag{9}$$

$$\varnothing_{st,i} = \frac{\pi(2i-3)}{T_s} + \varnothing_{ss,1} \tag{10}$$

where T_s represents total stator slots, while $\varnothing_{ss}$ and $\varnothing_{st}$ represent the center of slot and teeth, respectively.

4. Analytical derivation of stator and rotor MMF with reference to modified slot conductors

4.1. Analytical expression of stator MMF

In induction motors, the primary goal of windings is to produce a rotating MMF. Therefore, the windings of induction motors are designed in such a way that the induced EMF or produced MMF consists predominantly of the space-fundamental sinusoidal component. The phase windings of an induction motor are identical but distributed and shifted in space for $2\pi/3p$. The winding of each phase occupies one-third of the number of the total slots in an axial direction. The regions of stator winding within the slot known as slot conductors are of great importance because these are the portions of stator conductor that establish MMF and that are involved in the production of torque. The stator MMF is given as [29]

$$F_{ss}(t,\theta) = \sum\nolimits_{\mu} M_{\mu} \cos(\omega t - \mu p\theta)\ ;\ \mu = 6x+1;\ x = 0,\ \pm 1,\ \ \pm 2,\ \ \pm 3, \ldots. \tag{11}$$

where M_{μ} a harmonic term, representing the MMF amplitude of each harmonic p, refers to the total number of poles. It can be illustrated through the equation that besides the fundamental MMF there exist MMF waves of, i.e., *5p, 7p, 11p…n* pole pairs.

The stator winding is not a simple coil, but it consists of a number of identical coils that are placed in a distributive manner with certain angular coordinates as

$$W_{ss}(\theta) = \begin{cases} W_{ss} \cdot \left(1 - \dfrac{a}{2\pi}\right), & \theta_1 \le \theta \le \theta_2 \\ -W_{ss} \cdot \dfrac{a}{2\pi}, & \text{for other } \theta \end{cases} \tag{12}$$

W_{ss} is the slot conductors and a is a pitch.

4.2. Analytical expression of rotor MMF

The winding function of stator MMF is nothing else but MMF per unit current. In further text, the concept of stator MMF could be alternatively utilized to derive an expression for rotor MMF. Let us consider the rotor bars as a slot conductor and the end rings of the rotor will form a coil. Consequently, the combination of rotor bars with closed end ring forms coils at the rotor side:

$$W_{RR}(\theta) = \begin{cases} \left(1 - \frac{1}{R}\right), & -\frac{\pi}{R} \le \theta \le \frac{\pi}{R} \\ \frac{1}{R}, & \text{for other } \theta_R \end{cases} \tag{13}$$

The expression for rotor winding function can be derived by introducing Fourier series over the function in Eq. (13).

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{14}$$

$$W_{RR}(\theta_R) = \sum_{x=1}^{\infty} \frac{2}{x\pi} \sin\left(x\frac{\pi}{R}\right) \cos(x\theta_R) \tag{15}$$

Thus, following expression is achieved for rotor MMF:

$$F_{RR}(t, \theta_R) = \sum_{x=1}^{\infty} \frac{2}{x\pi} \sin\left(x\frac{\pi}{R}\right) \cos(x\theta_R) I_R \cos(s\omega t) \tag{16}$$

Apply product formula:

$$F_{RR1}(t, \theta_R) = \sum_{x=1}^{\infty} K_{\mu x} [\cos(s\omega t + x\theta_R) + \cos(s\omega t - x\theta_R)] \tag{17}$$

Referring to Eq. (11), it is clear that besides fundamental MMF there exist MMFs due to additional waves with $5p$, $7p$, $11p$,…, n. Consequently, the MMF of the rotor will also exit due to additional waves with $5p$, $7p$, $11p$. Further, **Figure 5** shows test rig designed to experimentally validate and capture MMF of induction motor. The test rig consists of 0.25 kW three-phase induction motor with 2 poles, 28 stator slots, and 30 rotor slots. While the spectrum of MMF is captured through PASPORT magnetic field sensor connected to PC through PASCO interface. The advantage of using PASPORT magnetic field sensor is that it offers very high sensitivity and is capable of monitoring the signal from −1000 to +1000 G as shown in **Figure 6**.

Figure 7 shows MMF spectrum against different harmonics, i.e., μ = 1,5,7,11, *and* 13 with corresponding μp = 2,10,14,22, *and* 26. It is more obvious that the additional harmonics waves will depend on the rotor bars and number of poles. Thus, besides the fundamental harmonic, rotor as well as stator MMF will exit for higher space harmonics. The abovementioned series of MMF due to μ harmonic term which could be defined as

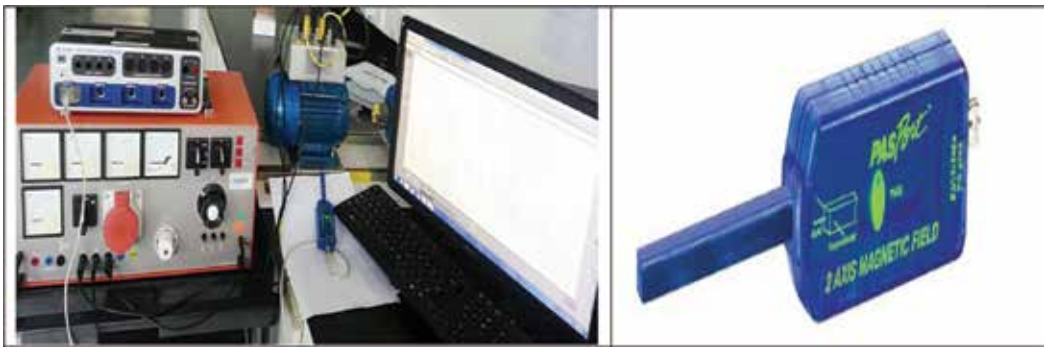


Figure 5. Test rig for measuring MMF.

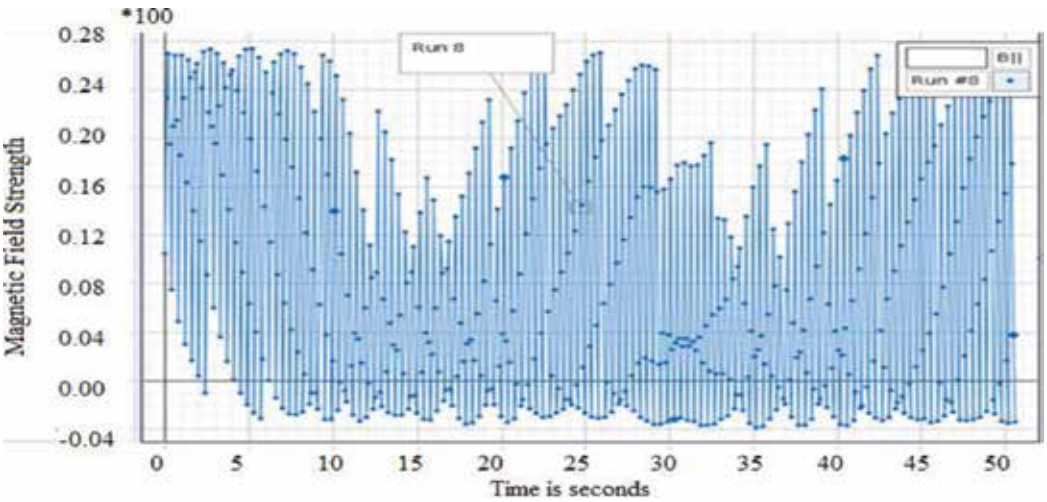


Figure 6. Spectrum of MMF two-axis magnetic field sensor.

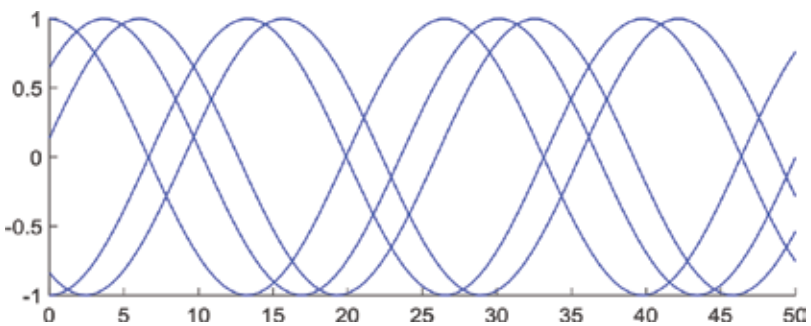


Figure 7. MMF at an instant ($p = 2$, $R = 30$, $\mu = 1$, 5, and 7).

$$\mu = 6u + 1;\ u = 0,\ \pm 1,\ \ \pm 2,\ \ \pm 3, \ldots.$$

The higher order harmonics are the consequence of the generation of rotor MMF. Due to analytical and experimental analysis, in this research work, an additional harmonic term "μ" is introduced to a rotor MMF. The rotor loops are $2\pi/R$ apart from each other with a phase shift of $\mu.p.2\pi/R$,; therefore, the frequency and the magnitude of the current will be same. The MMF in the consecutive rotor loop will be

$$F_{RR2}(t, \theta_R) = \sum_{x=1}^{\infty} K_{\mu x}\left[\cos\left(s\omega t + x\theta_R - (x + \mu p)\frac{2\pi}{R}\right) + \cos\left(s\omega t - x\theta_R + (x - \mu p)\frac{2\pi}{R}\right)\right] \tag{18}$$

The resultant MMF of the rotor will be the resultant of all the rotor loops along with all the harmonics.

$$F_R(t, \theta_R) = \sum_{j=0}^{R-1}\sum_{x=1}^{\infty} K_{\mu x}\left[\cos\left(s\omega t + x\theta_R - j.(x + \mu p)\frac{2\pi}{R}\right) + \cos\left(s\omega t - x\theta_R + j.(x - \mu p)\frac{2\pi}{R}\right)\right] \tag{19}$$

The harmonics in the rotor is due to fundamental and waves with 5*p*, 7*p*, 11*p*, …, *n* poles. Moreover, from Eq. (19) through inspection, it is found that more prominent MMF waves exist if $x = \pm\mu p$ as well as $x = \pm lR + \mu p$ or $x = \pm lR - \mu p$ where, $l = 1, 2, 3\ldots$. Suppose if it is stated that x should be a positive integer, then MMF waves only exist for higher order harmonics, i.e., $x = |\pm\ \mu p|$, $x = |\pm lR + \mu p|$ and $x = \pm lR - \mu p$. Thus, for each of these waves in addition to fundamental MMF $x = |\pm\ \mu p|$, there are rotor harmonics at $x = |\pm lR + \mu p|$ with respect to the waves with 5p, 7*p*, 11*p*, …*poles*. If these MMFs are observed from stator side, they will have the following expression:

$$F_{ss}(t, \theta_R) = F_{ssF}(t, \theta_s) + F_{ssRH}(t, \theta_s) + F_{ssRH}(t, \theta_s) \tag{20}$$

$$F_{ss}(t, \theta_s) = M_{\mu}\cos(\omega t - \mu p\theta_s) \tag{21}$$

$$F_{ssRH}(t, \theta_s) = M_{\mu}\cos\left(\left(1 - \frac{lR}{P}(1 - s)\right)\omega t + (lR - \mu p)\theta_s\right) \tag{22}$$

$$F_{ssRH}(t, \theta_s) = M_{\mu}\cos\left(\left(1 + \frac{lR}{P}(1 - s)\right)\omega t - (lR + \mu p)\theta_s\right) \tag{23}$$

To validate and justify aforementioned discussion, the spectrum of MMF is obtained against harmonics of (μ = *1, 5, and 7*) with respect to wave with 5*p*, 7*p*, 11*p*, …*poles*. From **Figure 8**, it is obvious that against fundamental wave with, i.e., μ = 1, there will be rotor harmonics of order $R - \mu P$, $R + \mu P$, $2R - \mu P$, and $2R + \mu P$, i.e., 8th, 32th, 58th, and 68th as shown in **Figure 8**(a). For μ = 5, the rotor harmonic of order 20th, 40th, 50th, and 70th are shown in **Figure 8**(b). Similarly, for μ = 7 and rotor harmonic are of order 16th, 44th, 46th, and 74th as shown in **Figure 8**(c). Similarly, the MMF at the stator side will exit for 5*p*, 7*p*, 11*p*,…. Further, this rotor harmonic will be useful to diagnose the faults.

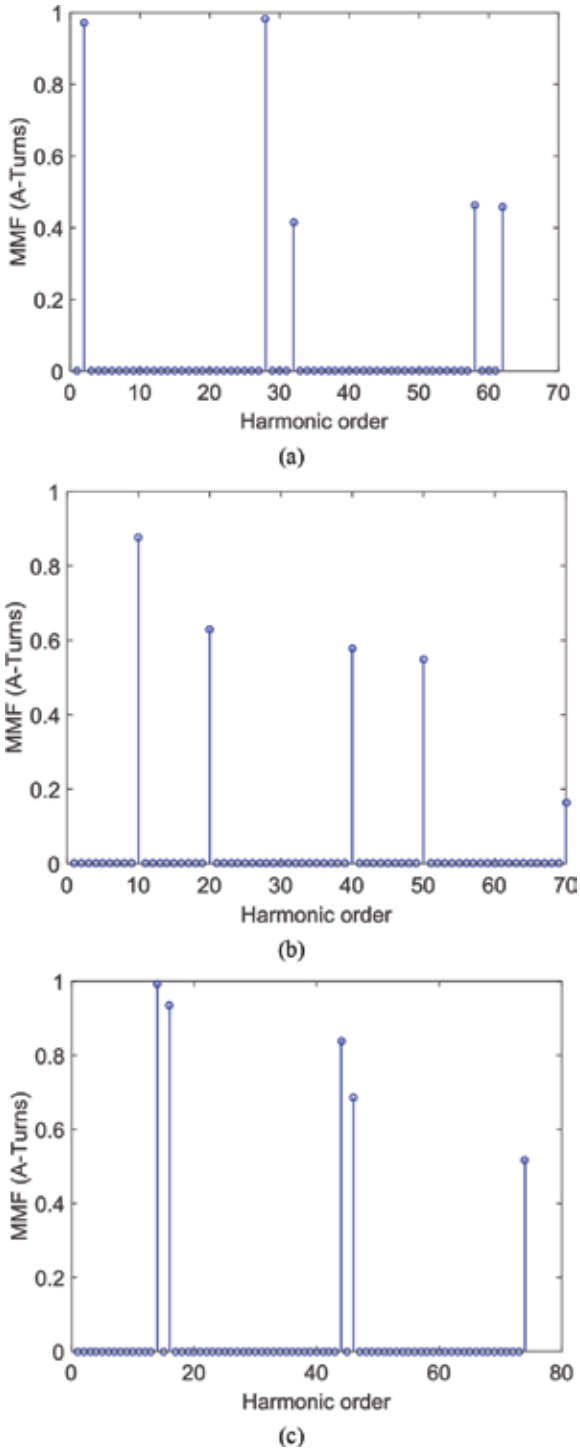


Figure 8. Harmonic order of MMF ($p = 2$, $R = 30$, $\mu = 1$, 5, and 7).

http://dx.doi.org/10.5772/67245

4.3. Analysis of unbalanced voltage and its effect on MMF

As mentioned earlier, unbalanced voltage supply is one of the major sources of stator winding failure. Under unbalanced voltage, the symmetry of induction motor is disturbed and the effect of unbalanced voltage is equivalent to negative sequence voltage. Motor acts as two different motors, one for the positive sequence and the other for the negative sequence. Each set of the sequence will produce corresponding balanced currents and the synthesis of the two sets of current is an actual current produced by the original unbalanced voltages. The response of a motor to the positive sequence voltage is the same as for normal voltage. However, the negative sequence currents set up a reverse MMF, so if a slip of a motor is (s) with respect to the positive sequence field, it will be (2-s) relative to the negative sequence MMF. Moreover, a motor behaves as two separate motors, one running at slip (s) with a terminal voltage while the other running with a slip of (2-s).

5. Fault diagnosis system through experimental results

For a reliable fault diagnostic method, it is essential to have a good understanding of the motor response for healthy and faulty conditions. To eliminate additional cost for monitoring scheme, it is important to find out the fault through current signature. The current signature can be easily measured and are most sensitive to any fault. Moreover, the cost of current monitoring is not imposed on the system because in every industry the current sensors are embedded. Thus, in proposed work, two methods are proposed to diagnose unbalanced voltage. In the first method, unbalanced voltage is diagnosed through an efficient MCSA technique. Due to the rapid complexity in industrial plants, it is inconceivable to continue human inspection to diagnose the faults. Consequently, to avoid human inspection, in addition to MCSA, AI-based ANN high performance fully automated method is proposed.

5.1. Diagnosis of unbalanced voltage through harmonic component

It is reported that MCSA is a noninvasive and an economically most suitable method to diagnose motor faults. Thus, the basis of MCSA is of fundamental importance to diagnose unbalanced voltage supply. Asymmetry in the power supply will generate negative sequence current that will produce a backward rotating field at double slip frequency with respect to the forward rotating rotor. The MMF that is caused by both the fields will induce voltage and current at the stator side, and the effect of this current will be more prominent at newly introduced rotor harmonic frequency given as

$$f_{RH} = \left[1/2\left(k\frac{2N_r}{P}\right)(1-s) \pm 1\right]F_s,\ k = 1, 2, 3\ldots\ldots\ldots \tag{24}$$

where, N_r and p represent rotor bars and poles, respectively, while s and F_s are the slip and applied frequency, respectively. It can be analyzed through modified expression of rotor harmonic that spectral content of current will be more prominent at these frequencies.

5.2. Experimental diagnosis of unbalanced voltage

To validate the proposed methods at an incipient stage, experimental data and justification are required. For this purpose, a number of experiments are carried on the induction motor. The motor is assembled and designed in electrical machine lab to study the behavior of the motor under different degree of unbalanced voltage in a single phase.

The block diagram and a testbed are shown in **Figures 9** and **10**, which are composed of a three-phase squirrel cage induction motor connected in star layout with, two pole pairs, variable adjustable voltage. The variation in the voltage supply is achieved through Lab-Voltage variable Power supply model 8821-2A. For data acquisition, FLUKE interface and PC are used. **Table 1** shows the rated specification of the motor that is used for an experiment, while **Table 2** shows the percentage unbalanced voltage that is introduced through variable Power supply model 8821-2A.

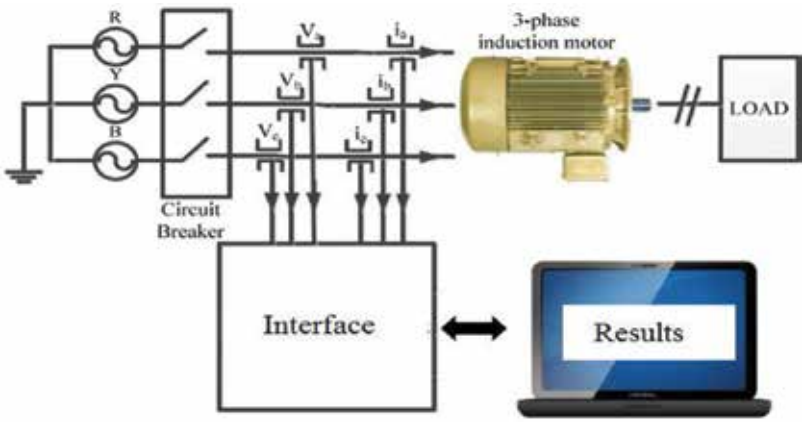


Figure 9. Test rig.

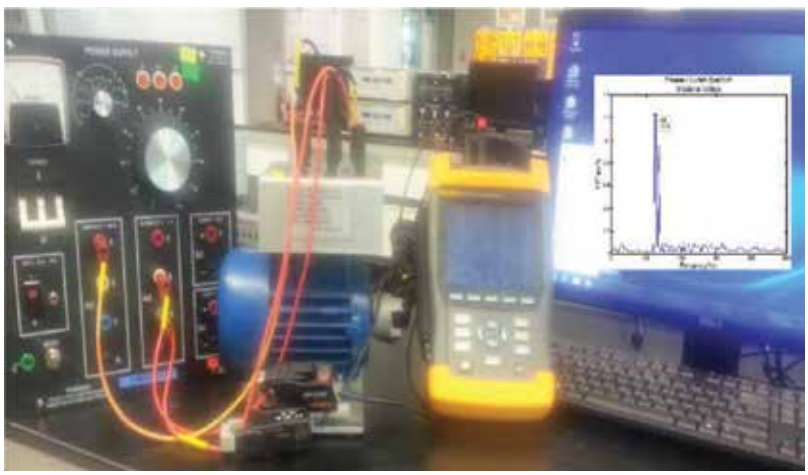


Figure 10. Experimental setup.

http://dx.doi.org/10.5772/67245

Motor feature	
Motor type	Three-phase induction motor
Balanced line voltage (*V*)	415
Synchronous speed (rpm)	3000
Rated speed (rpm)	2700
Motor slip	0.1
Power (W)	250
Frequency (Hz)	50
Poles	2
Stator slots	24
Rotor slots	22

Table 1. Rated values of induction motor.

	Voltage phase_1 (V)	Voltage phase _2 (V)	Voltage phase _3 (V)	% total voltage unbalanced
Balanced voltage	415	415	415	0
Unbalanced voltage	415	415	385	7.5
Unbalanced voltage	415	415	355	15
Unbalanced voltage	415	415	335	21

Table 2. Asymmetry in voltage supply.

Figure 11 shows the experimental results of current signature of three-phase induction motor supplied with a perfectly balanced voltage source. Referring to analytical expression (24), from the current spectrum, it is obvious that the current harmonics components at 450 and 550 Hz are due to induced EMF. It can easily be seen that the new series of rotor harmonic are the most prominent in the current spectrum. Hence, these frequency components in the current spectrum can be selected as yardsticks to measure the effects of unbalanced voltage.

To investigate the effect of unbalanced voltage, the percentage unbalanced in a single phase of an induction motor is varied through Lab-Voltage variable Power supply model 8821-2A. Further, the system response is analyzed through a new series of rotor harmonic. It is obvious from **Figures 12–14** that show that with the variation in the voltage supply there is a significant variation in the amplitude of the harmonic component at 450 and 550 Hz. It is also apparent that the amplitude of these harmonic is directly proportional to the degree of unbalanced voltage. Thus, the increase in the degree of unbalanced voltage will significantly increase the harmonic components at 45 and 550 Hz. Unbalanced voltage supply has drastic consequences on the motor operation because it can lead the current 6–10 times higher than the nominal rate. The increase in current will be much severed for the winding because even a 10% increase in temperature damages half of the winding insulation and results in short circuit. Thus, the proposed method will be a more appropriate and noninvasive technique to diagnose unbalanced voltage supply at in incipient stage and avoid catastrophic failure.

5.3. Results and discussions

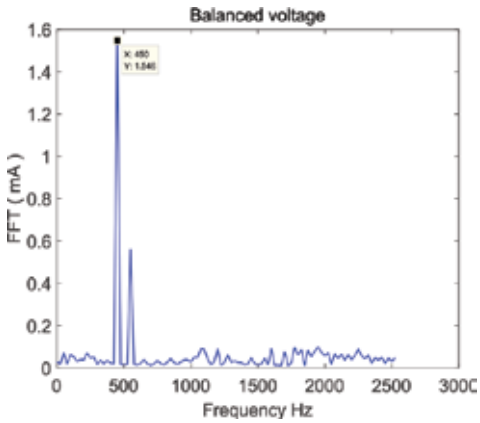


Figure 11. Balanced voltage.

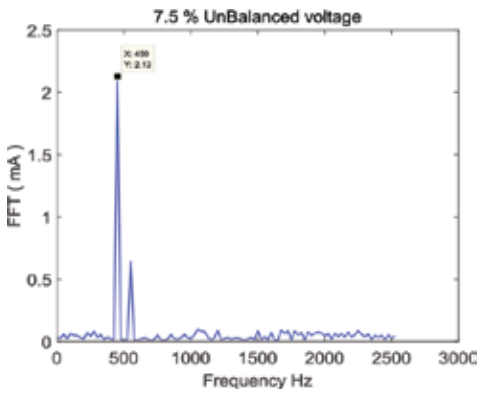


Figure 12. 7.5% Unbalanced voltage.

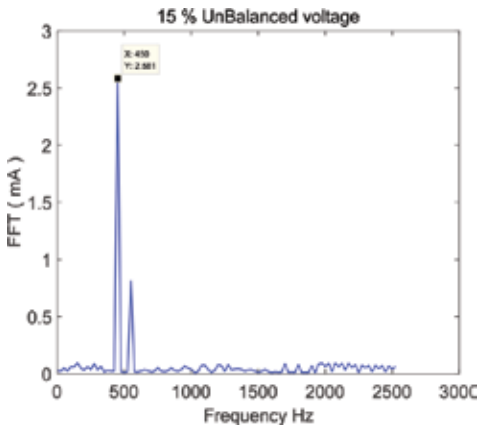


Figure 13. 15% Unbalanced voltage.

http://dx.doi.org/10.5772/67245

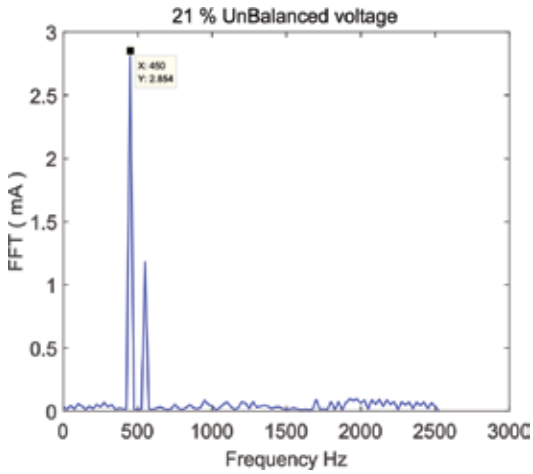


Figure 14. 21% Unbalanced voltage.

6. Fully automated detection of unbalanced voltage

6.1. ANN classifier design

The recognition of the pattern has been a key factor in the growth of fault diagnosis methods. Over the last few decades, automated techniques adopting the ANN have attracted the researchers due to excellent innovation in various industrial applications such as aerospace, pneumatic, chemical, and renewable energy. Similarly, in an electrical field, a major part of the research has been conducted on parameters such as identification in control systems, power system, and particularly detection of induction motor faults. The ANN is most feasible and reliable approach because these techniques are less model dependent of the process and in their generalization capacity. Moreover, ANN provides interpretation of real-time tracking by accommodating variation in the learning data. This efficiency is dependent on the choice of the data that are selected as inputs representing the defects. Hence, the ultimate obligation for effective implementation of ANN to diagnose the exact fault is the availability of relevant rich statistics that are identified as an input data for each fault. Therefore, it is compulsory that the input data of ANN should be meaningful indicators of the faults. Selection of such a data set is not an easy task. However, a convenient way is to select the parameter as a fault indicator that gives the most information about the fault and discard the rest [30]. In this work, feed-forward multilayer perceptron (MLP) ANN is used while the ANN is trained by the back-propagation (BP) algorithm.

6.2. Description of the proposed method

The choice of eccentricity harmonic detection, which increases the discernment between healthy and faulty motor, is of great interest that most of the researchers look for. From experimental results, most of the researchers confronted with the fact that the amplitude of third-order harmonic was considered alone distinction criterion between the healthy and

unbalanced voltage source, which is not sufficient owing the unbalanced condition because the third-order harmonic also exits for a perfect balanced source. In proposed work, the main focus is to use ANN technique in such a way that it is able to identify and declare healthy or faulty condition after reception of each data set. To acquire the desired results, certain stages must be carried out such as relevant input variables, output variables, and layers and neurons in each layer as shown in **Figure 15**.

The proposed method is based on the current signature of an induction motor. From the current signature, two indicators of voltage unbalanced are extracted, i.e., (I-1, I-2) based on new series of rotor harmonic (Eq. (20)). As shown in **Figure 16**, these indicators are fed to the ANN for automatic fault detection. The computation of this input is performed as illustrated in **Figure 16**. Initially, fast Fourier transformer (FFT) is performed at line currents to extract the magnitude and the phase angle of each line current. After that, the magnitude and phases of certain harmonic components are extracted and fed to ANN. Accordingly, the used multilayer feed-forward neural network with two input and four outputs is depicted in **Figure 16** with:

I-1: The magnitude of new series of rotor harmonic.

I-2: The phase angle of new series of rotor harmonic.

The neural network approach to diagnose unbalanced voltage is carried out in two phases, i.e., training and testing. In the first phase known as training phase, the ANN is automated trained to capture the underlying relations among inputs and desired outputs. While in the second phase the results are computed through ANN testing, the data set used for testing is random,

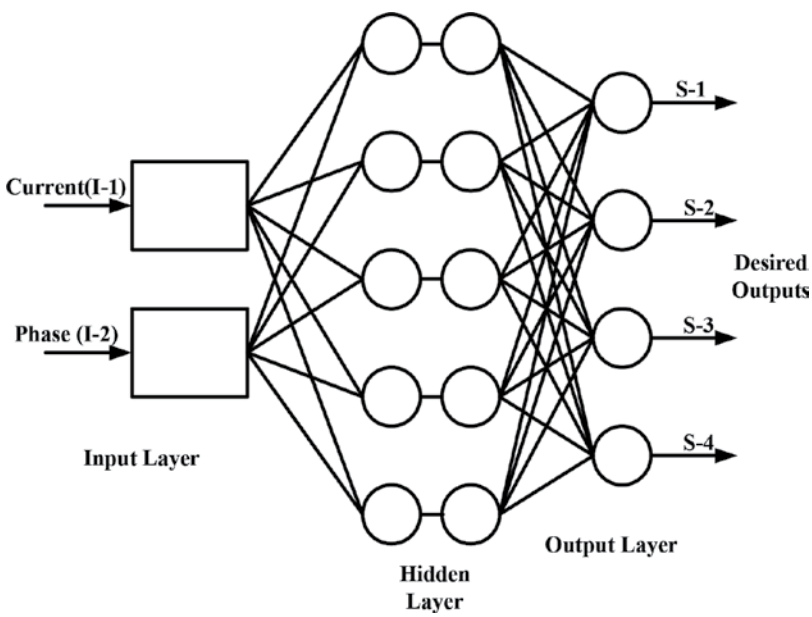


Figure 15. ANN architecture.

http://dx.doi.org/10.5772/67245

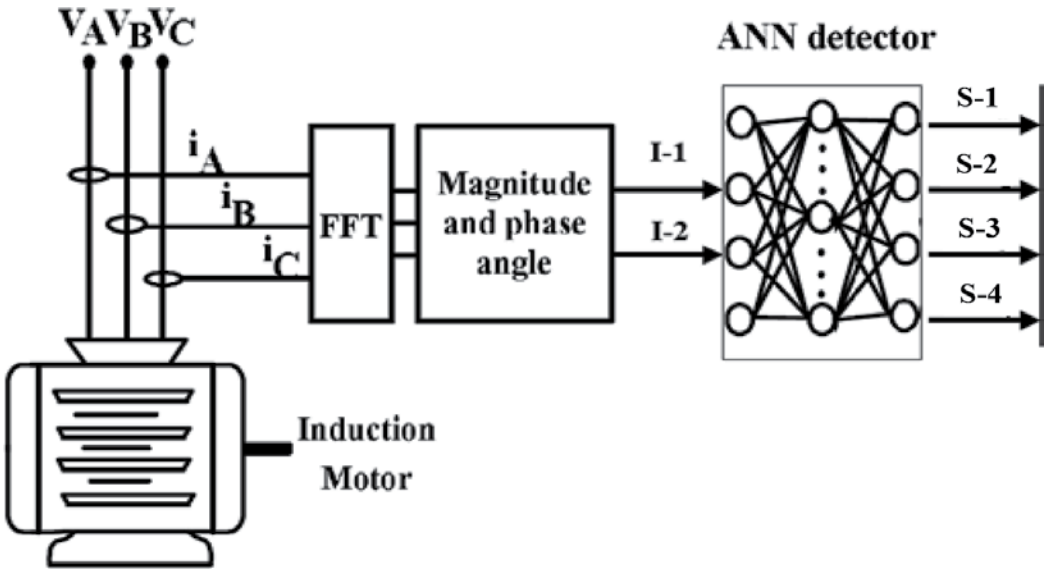


Figure 16. Model of proposed ANN multilayer feed-forward neural network.

which is not used before to check the efficiency of ANN. Note that in this proposed work, experimental data are collected for the testbed designed in electrical machine lab. The variation in the voltage supply is achieved through Lab-Voltage variable Power supply model 8821-2A.

6.3. Experimental results of the proposed ANN technique

6.3.1. Testbed description

To validate the proposed methods, experiments are carried on the induction motor. The testbed is shown in **Figure 10**, and its parameters are listed in **Table 1**. The variation in the voltage supply is achieved through Lab-Voltage variable Power supply model 8821-2A. For data acquisition, FLUKE interface and Intel (R) Xeon (R) computer having CPU E5-1650V2@3.50GHz, 64 GB RAM, and 64-bit operating system are used. Furthermore, for data acquisition, these analogy sensors are connected to PC through FLUKE interface. A total of 2040 samples of data for each class from FLUKE interface are interpreted by MATLAB (R2015a) software in order to compute the desired harmonics and the phase shift of each phase. This task is carried out through developed algorithm and the achievement of the fault diagnosis task by the NN.

In this research, different sequences of tests are conducted on the motor. The results of these tests are the basis to carry-out the next two phases of fault detection technique, i.e., training and the test phase, in order to experimentally validate the performance of the NN. The tests are conducted by configuring degree of unbalanced voltage in phase-C, for the following variations:

1. All the three phases are perfectly balanced.
2. Phase-C with 7.5% of unbalanced voltage while other two phases with nominal voltage.
3. Phase-C with 15% of unbalanced voltage while other two phases with nominal voltage.
4. Phase-C with 21% of unbalanced voltage while other two phases with nominal voltage.

6.3.2. Experimental characteristics of new series of harmonic

The plot in **Figures 17–19** illustrates the experimental characteristics in a new series of rotor harmonic for unbalanced voltage, respectively, with increasing degree of the unbalanced voltage of 7.5, 15, and 21%. It is obvious from the experimental computed result that there is a direct impact of unbalanced voltage on the amplitude of new rotor harmonics. With increasing unbalanced voltage, there is a direct increase in the amplitude of new series of harmonics in phase-A and phase-B, while there is a decrease in the amplitude of phase-C, which reflects that the unbalanced is in phase-C. Despite the difference in values, the evolution in the amplitude of plotted curves against increasing percentage of unbalanced voltage is an indicator to the permutation in the increasing degree of unbalanced voltage.

6.3.3. Training results

A number of input and output data set are processed during the learning phase. Thus, a series of training data set is fed to the ANN. The training data is extracted from the current signature and the targeted input database (I-i) with corresponding output database (S-i). The input data set of training is composed of different degree of unbalanced voltage. Each training data set is represented by input and output vectors, i.e., [I-1, I-2] = [S-1, S-2, S-3, S-4].

Total training input data set consists of 80 successive sequences of 20 examples to balanced voltage, 20 examples of 7.5% of unbalanced voltage, 20 examples of 15% of unbalanced

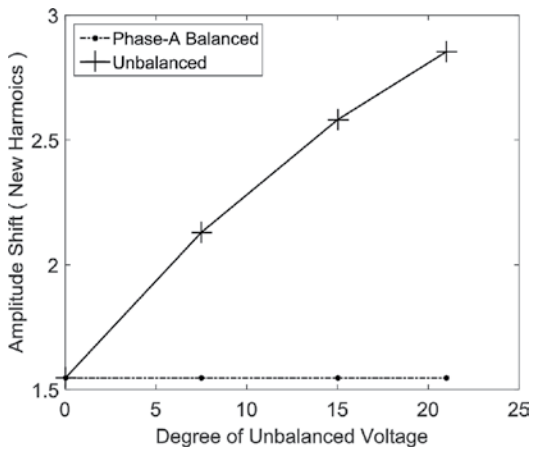


Figure 17. Characteristic of new harmonic on phase-A.

http://dx.doi.org/10.5772/67245

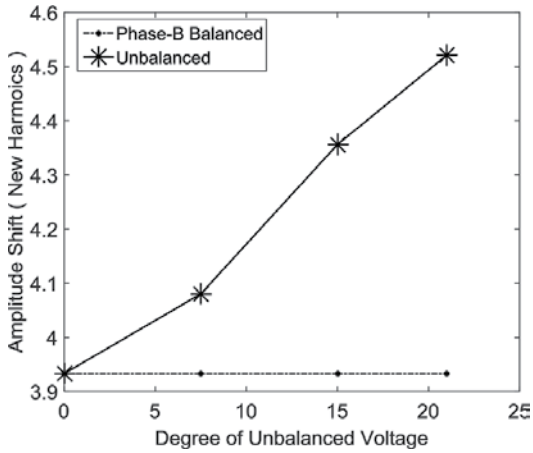


Figure 18. Characteristic of new harmonic on phase-B.

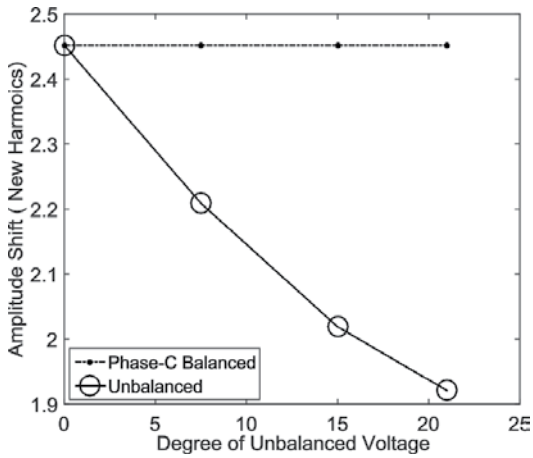


Figure 19. Characteristic of new harmonic on phase-C.

voltage, and 20 examples of 21% of unbalanced voltage in phase-C of an induction motor. The training output data are arranged in order to accomplish desired outputs (S-i) using a logical expression, i.e., (0 and 1) to identify the condition of the motor as follows:

S-1 = 1 All the three phases are perfectly balance; otherwise, S-1 = 0.

S-2 = 1 Phase-C with 7.5% of unbalanced voltage while other two phases with nominal voltage; otherwise, S-2 = 0.

S-3 = 1 Phase-C with 15% of unbalanced voltage while other two phases with nominal voltage; otherwise, S-3 = 0.

S-4 = 1 Phase-C with 21% of unbalanced voltage while other two phases with nominal voltage otherwise, S-4 = 0.

To elucidate the state, the outputs (S-1, S-2, S-3, S-4) are set to:

[0, 0, 0] All the three phases are perfectly balance.

[1, 0, 0] Phase-C with 7.5% of unbalanced voltage.

[0, 1, 0] Phase-C with 15% of unbalanced voltage.

[0, 0, 1] Phase-C with 21% of unbalanced voltage.

The total training input data set consists of 80 successive sequences of 20 examples to balanced voltage, 20 examples of 7.5% of unbalanced voltage, 20 examples of 15% of unbalanced voltage, and 20 examples of 21% of unbalanced voltage in phase-C of an induction motor. Using the NN structure with two inputs, four outputs, and 10 hidden neurons, as shown in **Figure 16**, the classifier is trained on the experimental data shown in **Figures 17–19**. It is obvious that satisfactory training results are achieved. From **Figure 20**, it can be perceived that mean squared error (MSE) reached to 4.63×10^{-09} after 1000 epochs. The corresponding MSE shows that successful classification is accomplished with low errors.

6.3.4. Test results

The objective of test results is to evaluate the feasibility and accuracy of the NN. Thus, to test the performance of proposed method, random and successive range of data is fed to ANN classifier with 20 examples of each case. From the MSE results, it can be analyzed that error is very low, and this is the indication that the proposed ANN model has correctly classified all the classes. Usually, the classifier performance is evaluated by estimating its generalization error. However, in many cases the low value of error is not enough to judge and ensure that the features were correctly learnt because too low error can also refer that the network has over fitted the data [31].

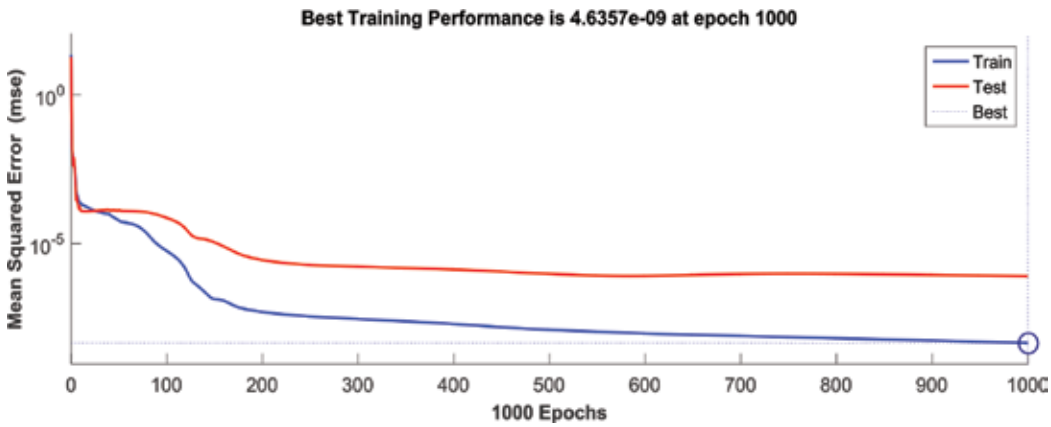


Figure 20. Performance of proposed ANN.

http://dx.doi.org/10.5772/67245

To properly classify the network performance Receiver Operating Characteristic (ROC) and confusion matrix are presented in this work. The ROC curve is a two-dimensional measurement that represents the classifier performance rate. It expresses the probability rate of corrected samples classified versus misclassified ones. In practice, ROC curve shows performance of proposed method through area under the curve. The performance

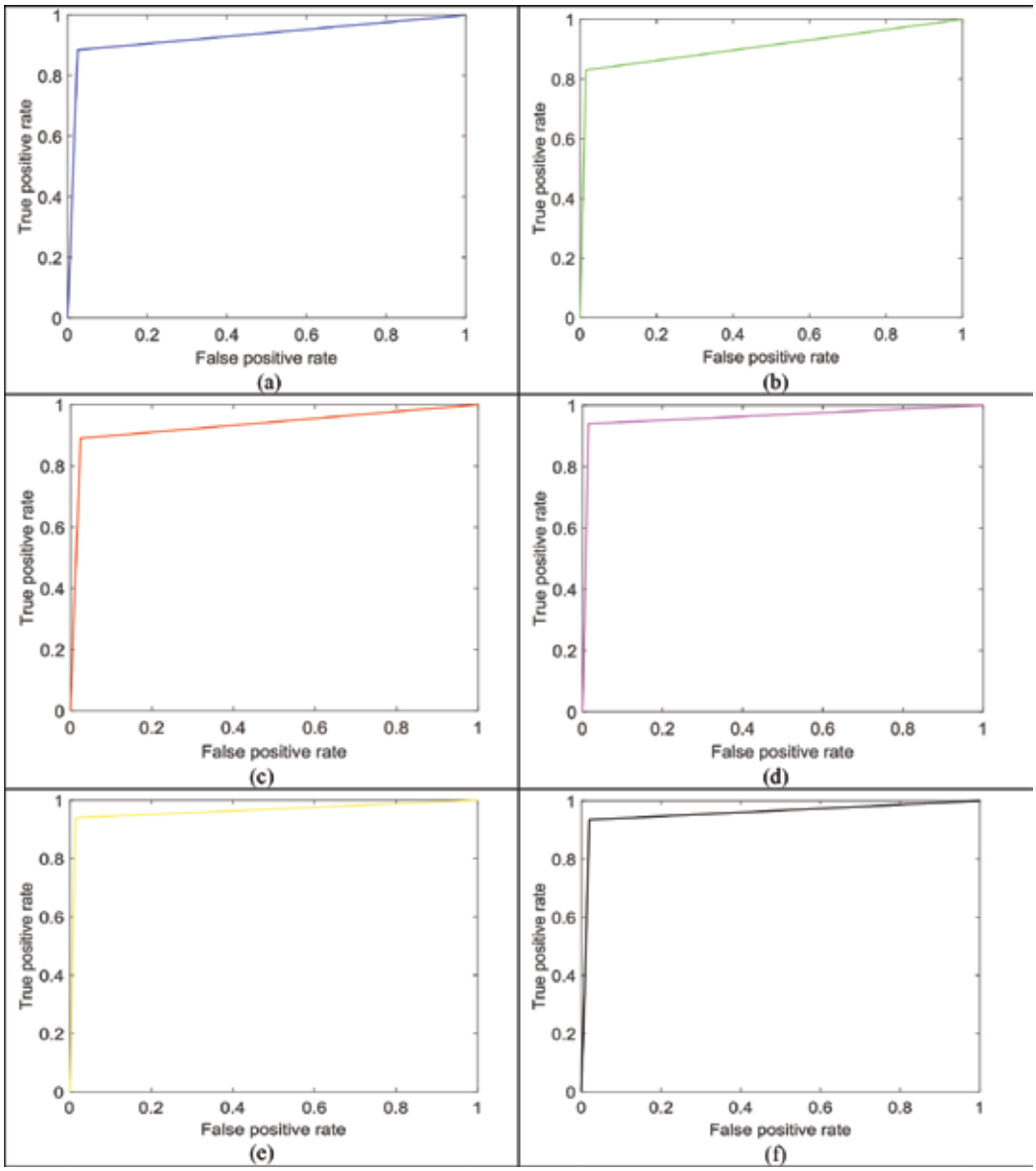


Figure 21. ROC curve w.r.t. (a) balanced-unbalanced 7.5%, (b) balanced-unbalanced 15%, (c) balanced-unbalanced 21%, (d) unbalanced 7.5%-unbalanced 15%, (e) unbalanced 7.5%-unbalanced 21%, (f) unbalanced 15%-unbalanced 21%.

of the network depends on the trade-off between robustness and sensitivity. The robustness and sensitivity of proposed techniques can be discovered through a series of experiments. ROC curve is carried out to detect and evaluate the proposed methodology that how efficiently propose technique diagnoses the level of unbalanced voltage supply. The ROC curve can be evaluated for each class to extract the performance against that particular class. From **Figure 21**, the ROC graph shows the trade-off between the probabilities of true positives rate (tpr) versus the probability of false positives rate (fpr). From ROC curve of each class, it can be seen that for all the classes the curve is closer to the top left corner which demonstrate that area under the curve is higher and the proposed method has efficiently diagnosed the level of unbalanced voltage supply. Mathematically, tpr and fpr can be expressed as

$$tpr = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \tag{25}$$

$$fpr = \frac{\text{false positives}}{\text{false positives} + \text{true negatives}} \tag{26}$$

Further, confusion matrix is constructed of each class to report the performance of a classifier. In **Table 3**, a confusion matrix is presented with dimension defined by the number and the type of classes. For clarification, the headings pointed out in confusion matrix of **Table 3** are defined as follows: 0 ≡ balanced voltage supply, 1 ≡ 7.5% unbalanced voltage supply, 2 ≡ 15% unbalanced

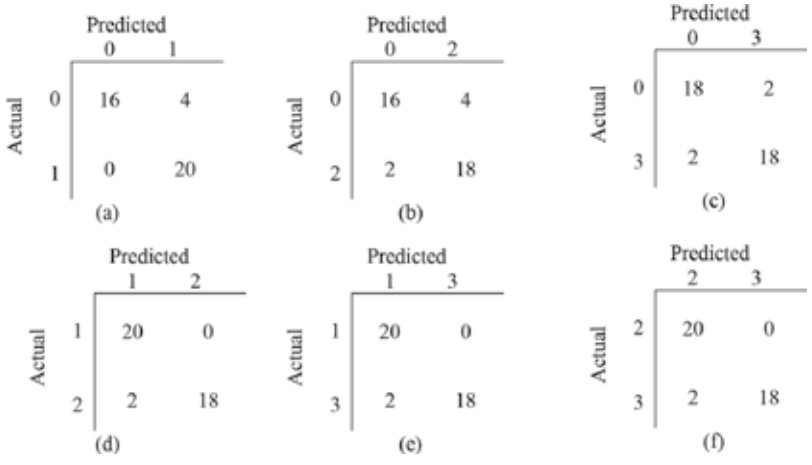


Table 3. Confusion matrix of (a) balanced-unbalanced 7.5%, (b) balanced-unbalanced 15%, (c) balanced-unbalanced 21%, (d) unbalanced 7.5%-unbalanced 15%, (e) unbalanced 7.5%-unbalanced 21%, (f) unbalanced 15%-unbalanced 21%.

http://dx.doi.org/10.5772/67245

voltage supply, and 3 ≡ 21% unbalanced voltage supply. The percentage of proper or correct identification of unbalanced voltage is represented in the major diagonal, while the out of major diagonal represents the errors (means the confusion) among the various classes. The results of confusion matrix demonstrate high degree of accuracy that shows that the classification associated with the use of proposed method is reliable to diagnose and distinguish between different levels of unbalanced voltage supply. The overall accuracy of proposed model is 85%, which is a big achievement.

The synthesis of the proposed work helps to conclude the feasibility and efficiency through the accuracy of the result that how close the ANN results are to the actual results. In this work, it is illustrated through experimental and classifier work that how accurate the proposed method is in locating the degree of unbalanced voltage. Moreover, the proposed method is robust and can be suggested to demonstrate the health of any motor.

7. Conclusion and future directions

As in recent era, the majority of induction motor fault diagnosis techniques are essentially noninvasive (online) to ensure continue production, reliability, and consequently. Therefore, in this chapter, a comprehensive analysis of noninvasive techniques is carried out, which are used to diagnose stator winding failure. Further, this chapter presents derivation by utilizing winding function approach to find out the conductor segments responsible for the generation of MMF. The knowledge of induced voltage at stator winding through MMF helps to propose a new noninvasive regime in order to diagnose unbalanced voltage. Thus, a new series of rotor harmonics are introduced and one can easily observe the performance through a graph that the amplitude of higher order harmonics increases against the various level of unbalanced voltage. Due to the rapid complexity in industrial plants, it is inconceivable to continue human inspection to diagnose the faults. Thus, to avoid human inspection, in additional to new series of rotor harmonic, a fully automatic method based on a neural network is proposed to diagnose unbalanced voltage at an incipient stage. The diagnostic process of ANN is fully automated; it can be observed through the pattern or the trend of ANN graphs that there is a significant rise in the new series of current harmonic. Further, through ROC curve, confusion matrix and overall accuracy of proposed method show that the novel features extraction are more reliable for diagnosing the fault. The classification tasks of proposed automated noninvasive method not only diagnoses unbalanced voltage but it also distinguishes between various levels of unbalanced voltage, which reflects the effectiveness and reliability of the proposed diagnosis process. The scope of this research can readily be extended to identify and locate the exact unbalanced voltage phase through a current spectrum. In proposed method, the features are extracted through the current harmonics, while the research can be extended to extract the unbalanced features through phase shift.

Author details

Muhammad Aman Sheikh[1]*, Nursyarizal Mohd Nor[1], Taib Ibrahim[1], Sheikh Tahir Bakhsh[2], Muhammad Irfan[3] and Hanita Binti Daud[4]

*Address all correspondence to: amanatd88@yahoo.com

1 Department of Electrical & Electronic Engineering, Universiti Teknologi Petronas, Bandar Seri Iskandar, Malaysia

2 Computer Science Department, Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia

3 Electrical Engineering Department, Faculty of Engineering, Najran University, Kingdom of Saudi Arabia

4 Department of Fundamental & Applied Sciences, Universiti Teknologi Petronas, Bandar Seri Iskandar, Malaysia

References

[1] M. Irfan, N. Saad, R. Ibrahim, V. S. Asirvadam, N. Hung, and M. A. Magzoub, "Analysis of bearing surface roughness defects in induction motors," *Journal of Failure Analysis and Prevention*, vol. 15, pp. 730–736, 2015.

[2] M. Irfan, N. Saad, R. Ibrahim, and V. Asirvadam, "An on-line condition monitoring system for induction motors via instantaneous power analysis," *Journal of Mechanical Science and Technology,* vol. 29, pp. 1483–1492, 2015.

[3] W. Thomson, "A review of on-line condition monitoring techniques for three-phase squirrel-cage induction motors-past present and future," in *Keynote address at IEEE symposium on diagnostics for electrical machines, power electronics and drives, Gijon, Spain*, 1999, pp. 3–18.

[4] W. T. Thomson and M. Fenger, "Current signature analysis to detect induction motor faults," *Industry Applications Magazine, IEEE*, vol. 7, pp. 26–34, 2001.

[5] S. Nandi, S. Ahmed, and H. A. Toliyat, "Detection of rotor slot and other eccentricity related harmonics in a three phase induction motor with different rotor cages," *Energy Conversion, IEEE Transactions on*, vol. 16, pp. 253–260, 2001.

[6] H. A. Toliyat and N. A. Al-Nuaim, "Simulation and detection of dynamic air-gap eccentricity in salient pole synchronous machines," in *Industry Applications Conference, 1997. Thirty-Second IAS Annual Meeting, IAS'97., Conference Record of the 1997 IEEE*, 1997, pp. 255–262.

[7] M. E. H. Benbouzid, M. Vieira, and C. Theys, "Induction motors' faults detection and localization using stator current advanced signal processing techniques," *IEEE Transactions on Power Electronics*, vol. 14, pp. 14–22, 1999.

[8] A. Siddique, G. Yadava, and B. Singh, "A review of stator fault monitoring techniques of induction motors," *Energy conversion, IEEE Transactions on*, vol. 20, pp. 106–114, 2005.

[9] G. Acosta, C. Verucchi, and E. Gelso, "A current monitoring system for diagnosing electrical failures in induction motors," *Mechanical Systems and Signal Processing*, vol. 20, pp. 953–965, 2006.

[10] J. Jee-Hoon, J.-J. Lee, and K. Bong-Hwan, "Online Diagnosis of Induction Motors Using MCSA," *IEEE Transactions on Industrial Electronics*, vol. 53, pp. 1842–1852, 2006.

[11] S. Nandi and H. A. Toliyat, "Novel frequency domain based technique to detect incipient stator inter-turn faults in induction machines," in *Industry Applications Conference, 2000. Conference Record of the 2000 IEEE*, 2000, pp. 367–374.

[12] M. Eltabach, J. Antoni, and M. Najjar, "Quantitative analysis of noninvasive diagnostic procedures for induction motor drives," *Mechanical Systems and Signal Processing*, vol. 21, pp. 2838–2856, 2007.

[13] N. Arthur and J. Penman, "Induction machine condition monitoring with higher order spectra," *IEEE Transactions on Industrial Electronics*, vol. 47, pp. 1031–1041, 2000.

[14] R. Isermann, "Model-based fault-detection and diagnosis–status and applications," *Annual Reviews in Control*, vol. 29, pp. 71–85, 2005.

[15] V. Uraikul, C. W. Chan, and P. Tontiwachwuthikul, "Artificial intelligence for monitoring and supervisory control of process systems," *Engineering Applications of Artificial Intelligence*, vol. 20, pp. 115–131, 2007.

[16] M. A. Awadallah and M. M. Morcos, "Application of AI tools in fault diagnosis of electrical machines and drives-an overview," *IEEE Transactions on Energy Conversion*, vol. 18, pp. 245–251, 2003.

[17] F. Filippetti, G. Franceschini, C. Tassoni, and P. Vas, "AI techniques in induction machines diagnosis including the speed ripple effect," *IEEE Transactions on Industry Applications*, vol. 34, pp. 98–108, 1998.

[18] A. Siddique, G. Yadava, and B. Singh, "Applications of artificial intelligence techniques for induction machine stator fault diagnostics: review," in *Diagnostics for Electric Machines, Power Electronics and Drives, 2003. SDEMPED 2003. 4th IEEE International Symposium on*, 2003, pp. 29–34.

[19] J. S. Hsu, "Monitoring of defects in induction motors through air-gap torque observation," *IEEE Transactions on Industry Applications*, vol. 31, pp. 1016–1021, 1995.

[20] M. Sahraoui, S. Zouzou, A. Ghoggal, and S. Guedidi, "A new method to detect inter-turn short-circuit in induction motors," in *2010 XIX International Conference on Electrical Machines (ICEM)*, 2010, pp. 1–6.

[21] S. M. A. Cruz and A. J. M. Cardoso, "Multiple reference frames theory: a new method for the diagnosis of stator faults in three-phase induction motors," *IEEE Transactions on Energy Conversion*, vol. 20, pp. 611–619, 2005.

[22] A. Ukil, S. Chen, and A. Andenna, "Detection of stator short circuit faults in three-phase induction motors using motor current zero crossing instants," *Electric Power Systems Research*, vol. 81, pp. 1036–1044, 2011.

[23] N. Mehla and R. Dahiya, "An approach of condition monitoring of induction motor using MCSA," *International Journal of Systems Applications, Engineering and Development*, vol. 1, pp. 13–17, 2007.

[24] J. Cusido, L. Romeral, J. A. Ortega, J. A. Rosero, and A. Garcia Espinosa, "Fault detection in induction machines using power spectral density in wavelet decomposition," *IEEE Transactions on Industrial Electronics*, vol. 55, pp. 633–643, 2008.

[25] S. B. Lee, R. M. Tallam, and T. G. Habetler, "A robust, on-line turn-fault detection technique for induction machines based on monitoring the sequence component impedance matrix," *IEEE Transactions on Power Electronics*, vol. 18, pp. 865–872, 2003.

[26] F. Villada, D. R. Cadavid, J. E. Aedo, B. A. Benavides, and E. Velilla, "The negative sequence impedance as fault indicator in induction motors," *Revista Facultad de Ingeniería*, pp. 7–13, 2013.

[27] S. Nandi and H. A. Toliyat, "Novel frequency-domain-based technique to detect stator interturn faults in induction machines using stator-induced voltages after switch-off," *IEEE Transactions on Industry Applications*, vol. 38, pp. 101–109, 2002.

[28] S. Nandi, "Detection of Stator Faults in Induction Machines using Residual Saturation Harmonics," in *2005 IEEE International Conference on Electric Machines and Drives*, 2005, pp. 256–263.

[29] P. Lamim, R. Pederiva, and J. Brito, "Detection of stator winding faults in induction machines using an internal flux sensor," in *Diagnostics for Electric Machines, Power Electronics and Drives, 2007. SDEMPED 2007. IEEE International Symposium on*, 2007, pp. 432–437.

[30] M. Bouzid, N. Bellaaj, K. Jelassi, G. Champenois, and L. Signac, "Location of an inter turns short-circuit fault in the stator windings of induction motor by neural network," 2007.

[31] Demian, C., Cirrincione, G., and Capolino, G. A. (2002, November) "A neural approach for the fault diagnostics in induction machines," in *IECON 28th Annual Conference of the IEEE Industrial Electronics Society*, (Vol. 4, pp. 3372–3376). IEEE.

Chapter 12

Detection and Analysis of Petroleum Equipment Faults

Mihail Minescu and Ion Pana

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.68227

Abstract

The method fitness-for-service (FFS) provides the means by which the operator of a technical system can decide: It can continue to work safely, reducing of working parameters or stopping the equipment and reparation it. A case study concerning a natural gas pipeline is introduced. It brings some applicative aspects: the introduction of the failure probability as an indicator; the reduction of the degree of conservatism; a maintenance program based on pipeline intelligent gauge.

Keywords: fitness-for-service, pipeline, maintenance program

1. Introduction

While an equipment (pressure vessel, pipeline, tank, etc.) is pressurized and has a certain state of degradation, the operator must decide: Whether it can continue to work safely, reducing the working parameters or the equipment must be stopped and refurbished, avoiding injury of the personnel or other persons, and unexpected environmental accidents [1]. The method fitness-for-service (FFS) provides the means by which the operator can take these decisions based on reliable engineering knowledge.

The main factors that have to be considered when determining the applicability and limitations of a procedure for evaluating a pipeline by FFS are data available on pipeline, operation and maintenance history of the pipeline. For pipelines used to the transport of hydrocarbons standard, API 579 [2, 3] (whose assessment procedures are in turn based on the ASME B31G and the RSTRENG criteria [4]) recommends several levels of evaluation. Level 1—Evaluation procedures included in this level are aimed at securing conservative monitoring criteria that can be used with a minimum amount of information and inspections concerning the pipeline.

They can be implemented by the technical personnel of the user. Level 2—Evaluation procedures included in this level are designed to ensure a more detailed assessment, which leads to more precise results compared to Level 1 assessment. In this level, the information from inspections is consistent with those provided for Level 1, but using more laborious calculations for their interpretation. Level 2 assessment should normally be realized by technical staff with experience in assessments of this type. Level 3—Evaluation procedures included at this level are aimed at ensuring the accurate assessment, leading to more accurate results compared to Level 2 assessment. In this level, the most detailed information and recommended inspections of the pipeline are typically required, and analysis is based on numerical techniques such as finite element method or experimental techniques. It is expected that this level assessment to be carried out only by experts with proven experience and expertise in such evaluations. Many papers are devoted to this topic. Shekari et al. [5] have used FFS assessment methodology for process equipment to track and predict pitting corrosion. Pit density was modeled using a non-homogenous Poisson process and induction time for pit initiation is simulated as the realization of a Weibull process. The distributions of the operating pressure and the estimated burst pressure of the defected component are integrated with Monte Carlo simulations and first-order second-moment (FOSM) method to calculate the reliability index and probability of failure. Scano [6] has used FFS assessment for a pipeline connecting the boiler of a paper mill to the cogeneration turbine and the process headers. Because of the elevated number of in-service hours, an API 579-1 Level 3 assessment was required, and a FE shell model of the line was set up to evaluate plastic strain accumulation due to creep through a time-dependant inelastic analysis. The results of the assessment led to an estimate of 70,000 hours of residual life for the pipeline. Almeida et al. [7] have proposed a modeling of a pressure vessel under internal and external corrosion using the fitness-for-service (API 579). Non-destructive testing by ultrasound was used to obtain loss of thickness wall measurements for pressure vessel damaged and develop the modeling. The objective is to analyze and evaluate the values of maximum allowable working pressure (MAWP) provided by the fitness-for-service assessment using numerical thermal transient analysis using finite element. Janelle [8] has reviewed the technical basis for the fitness-for-service assessment procedures for general and local metal loss. Extensive validation of these procedures along with additional development was presented. The conclusions of the study are recommended as the best practices to be included in future versions of API 579. Adib-Ramezani et al. [9] have studied the notch stress intensity factor concept, and SINTAP structural integrity procedure is employed to assess gas pipelines integrity. The external longitudinal defects have been investigated via elastic-plastic finite element method results. The extracted evaluations are compared with the limit load analysis based on ASME B31G, modified ASME B31G, DNV RP-F101. The comparison among extracted safety factors exhibits that SINTAP predictions are located between lower and upper safety factor bounds. Ahammed [10] has used deterministic model to evaluate the remaining strength of corroded steel pipeline over time. This model evaluates the remaining strength of corroded steel pipeline over time. The model also can be used to evaluate the maximum allowable failure pressure of corroded pipelines. Ahammed [11] has developed previous calculation model. Because of the presence of nonlinearity in the limit state function and also of the presence of non-normal variables, the Level II advanced first-order second-moment iterative method is employed for carrying out reliability analyses. Li et al. [12] have used an original methodology for predicting corrosion remaining life of underground pipelines with a mechanically based

probabilistic model by taking effect of randomness into account in pipeline corrosion. The results show that the corrosion defect depth and radial corrosion rate are the key factors influencing pipeline failure probability and remaining life. Netto et al. [13] have studied the effect of external corrosion defects via a series of small-scale experiments and through a nonlinear numerical model based on the finite element method. The model was used to determine the burst pressure as a function of material and geometric parameters of different pipelines and defects. Teixeira et al. [14] have evaluated the reliability of pipelines with corrosion defects subjected to internal pressure using the first-order reliability method (FORM). The limit-state function is defined based on the results of a series of small-scale experiments and three-dimensional non-linear finite element analysis of the burst pressure of intact and corroded pipelines. Minescu and Pana [15] have demonstrated the equivalence of the results obtained with the assessments procedures API 579 and ASME B31G over a pipeline transport system. In Section2.1.1, there are revealed the novelty aspects of this work. In conclusion, it should be said that the FFS method is extensively applied in industry in various fields; the progress of theoretical and experimental applied methods has improved the application results; method deserves to be consistently applied in technologic systems.

2. Using fitness-for-service assessment method—case study

2.1. Characteristics of fitness-for-service method

A case study concerning a natural gas pipeline is introduced for example. In this evaluation, which is essentially a FFS method, different ways are suggested from those used in the standards DNV RP 101 [16], API 579 and ASME B31G.National Regulatory Authority for Energy (NRAE) from Romania supervises the activity of the transmission system operators (TSO) for petroleum products. Such the transporters are obliged to fulfill certain procedures. Effectiveness of these procedures is measured by several indicators as the number of defects per km of pipeline; accidents found during operation; accidents caused by third parties; complaints of customers, etc. [17]. Pipeline (both for liquid petroleum products and for the gaseous hydrocarbons) from Romania is inspected by pipeline intelligent gauge (PIG) technology. Appreciation of the failure limit of a pipeline (for the case when the pipeline destruction is possible because of corrosion defects) can be done in two ways [18, 19]:

a. It makes the difference between failure pressure and the pressure of the operating.

b. It makes the difference between the thickness of the resistance (usually 80% of the pipe wall thickness) and depth of corrosion of corresponding to the defect that was detected in the pipeline wall.

Mustaffa et al. [20] have achieved an excellent review over limit state methods. Recently, some authors have developed models for the limit state (based on similarity theory) including the geometrical parameters of the pipeline, geometrical characteristics of major defects and pipeline operating conditions. A good study over the subject has accomplished by Zecheru et al. [21]. This is another way of estimation of the limit state, which was added to the two methods above. Caleyo et al. [18] have used the first-order second-moment iterative reliability method, and the Monte Carlo integration technique and the first-order Taylor series expansion of the

limit state function (LSF) are used in order to estimate the probability of failure associated with each corrosion defect over time. De Leon and Macias [19] have studied the reliability of a pipeline using FFS method. Several degrees of spatial correlation are assumed for the corrosion in determined segments of a pipeline, and their effects on the global reliability are examined. The pipeline is assumed to be a series system. The failure mode is considered to be controlled by the stresses due to internal pressure and the presence of corrosion. Component reliability is calculated by first-order second-moment approximations. The defects identification and appreciation of their evolution in time are valuable if it ends with a maintenance program indicating when and where to intervene to repair the pipeline, before producing an unwanted incident. The application further described has the following enhancements:

a. Provide a maintenance program based on the information during the inspection. This program has implemented since 2008 in TSO main companies from Romania Transgaz SA and Conpet SA.

b. In the theoretical model further exposed, the operating pressure was considered in the place where the fault occurs. Considering the pressure at the defect position reduces the degree of conservatism of the evaluation method.

c. Based on geometrical parameters, characteristics of major defects and pipeline operating conditions can calculate the probability of pipeline failure. This indicator is better than traditional indicators used by NRAE as it includes measurement results PIG.

2.2. Theoretical model

The appreciation of the limit state of a pipeline can be made by several methods [12, 18, 20], from which, in this paper, it was used the difference between the failure pressure of the pipeline PF_i corresponding to the defect i and the pressure of operating PO_i corresponding to the position of this defect:

$$Z_i = PF_i - PO_i. \tag{1}$$

The pressure of failure has more computing methods [4, 15, 16], from which, in this paper, it was used for exemplification the RSTRENG[1] method:

$$PF_i = \frac{2 \cdot UTS \cdot t}{D}\left(1 - \frac{d_i}{t \cdot M_i}\right), \tag{2}$$

$$M_i = \sqrt{1 + 0.63\left(\frac{l_i}{D}\right)^2\left(\frac{D}{t_i}\right) - 0.0034\left(\frac{l_i}{D}\right)^4\left(\frac{D}{t_i}\right)^2}, \tag{3}$$

where UTS is the ultimate tensile strength of the material of the pipeline; t_i is the wall thickness of the pipeline at defect location; D is the outer diameter of the pipeline; d_i is the depth of the defect; M_i the bulging factor(Folias); l_i is the length of the defect. The pressure of operating at

[1]Remaining Strength of Corroded Pipe (RSTRENG) assessment procedure, www.rstreng.com.

the defect position PO_i was calculated considering a linear variation of the pressure along the pipeline:

$$PO_i = PO_{start} - \frac{PO_{start} - PO_{end}}{L_p} L_i, \tag{4}$$

where PO_{start} is the operating pressure at the inlet of the pipeline; PO_{end} is the pressure at the outlet of the pipeline; L_p is the length of the pipeline; L_i is the distance from the beginning of the pipe at the location of the defect i. The values used in the relations above are l_i, d_i, t_i, L_i from the results of the PIG inspection; D, PO, UTS probabilistic variables (the mean value is known and the value of standard deviation is based on statistical studies [15, 21–23] **Table 4**). To calculate the probability of failure FP_i of the defect i, the Monte Carlo method [5, 18] was used. If the difference expressed by Eq. (1) is positive, the situation is favorable and the pipeline does not fail. If the difference is less or equal to zero, the pipeline fails. We note the number of attempts for $Z \leq 0$ with n_d. The probability of failure expressed for a number of N tests performed is as follows:

$$FP_i = \frac{n_d}{N}. \tag{5}$$

For a pipeline with a number of defects n, it is considered a system in series with n critical elements, and the probability of failure FP is [19]:

$$FP = 1 - \prod_{i=1}^{n} (1 - FP_i). \tag{6}$$

The variation of the size of defect over time (the time is denoted with T) was calculated with the relations:

$$l_i(T) = l_i(T_0) - V_{a,i}(T - T_0), \tag{7}$$

$$d_i(T) = d_i(T_0) - V_{r,i}(T - T_0), \tag{8}$$

where T_0 is the time of inspection of the pipeline. We considered the values of the corrosion rates (at each defect i) in the axial direction $V_{a,i}$ and in the radial direction $V_{r,i}$, and their values were determined at the time of the inspection (and constant further considered):

$$V_{a,i} = \frac{l_i(T_0)}{T_0}, \tag{9}$$

$$V_{r,i} = \frac{d_i(T_0)}{T_0}. \tag{10}$$

2.3. Results obtained

For example, it was used a steel pipeline (52 × SR 11082 material and 57km length) located between cities Constanta and Ploiesti. The example is extracted from research work [23]. The

pipeline was inspected using an ultrasound method [the usual methods of inspection there are magnetic flux and ultrasound and the device is named pipeline intelligent gauge (PIG)], **Figure 1**. We used for the inspection a20″ Ultrasonic Intelligent Pig, Korsonic 324, with the following specifications: PIG diameter 350 mm; body length 850 mm; overall pig length 950 mm; temperature max. for PIG 65°C; pressure max. 50 barg; min. bend radius 3 × Internal diameter; transducer frequency 0.5 MHz; transducer focus plane; min. measurable thk 3 mm; max. measurable thk 0.7 m; inspection sensitivity ±0.1 mm; repetition rate 2300 kHz; inspection speed Max. 5 m/min; max. inspection capacity 120 h; axial sampling distance min. 3 mm; circumferential resolution 5.5 mm. The ultrasonic signal is induced directly in the wall to be inspected, EMAT technology. It notes that the procedure for determining defects of the pipelines uses three-dimensional images that are offered to users in the form of Excel files, **Table 1**. These images were obtained over the last 20 years with a precision increasingly better. The instrument measures the thickness of the pipeline in a network of points, **Figure 2**. The image is reported as an Excel file. An example is shown in **Table 1**. As the beneficiary of the contract imposed certain conditions of confidentiality, they have been used data from a pipeline segment of 8622 m. So the probability of failure calculation refers only to this segment and not to the entire pipeline. The total number of defects found was 56.824. These can be classified after the geometrical characteristics (**Table 2**) and the cause that determined the defect: manufacturing, construction, corrosion, mechanical damage and repair (**Table 3**, column 10).

The defects characteristics were included in a data matrix size (56,824; 10) each row representing a defect, **Table 3**. The significance of columns of the data matrix is as follows: the distance at which the welds are located on the pipeline segment measured from the start of the pipeline

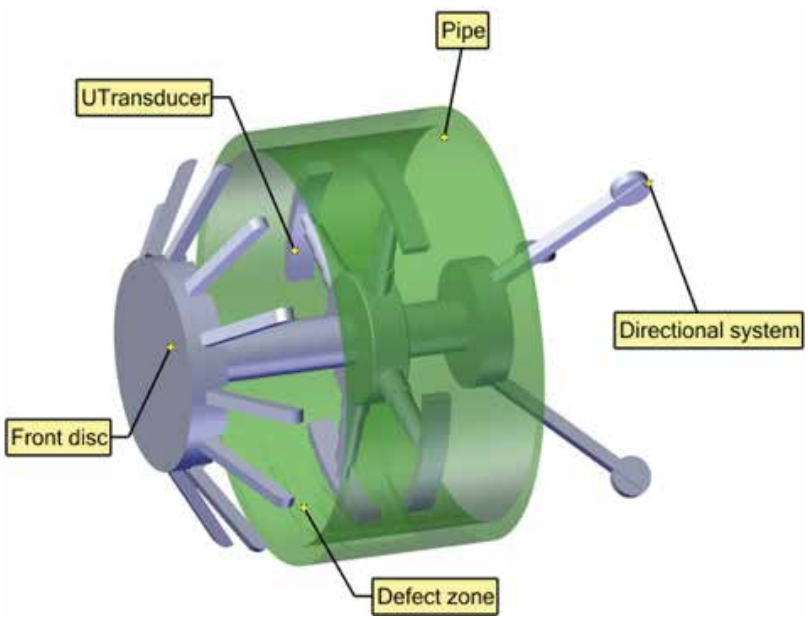


Figure 1. Pipeline intelligent gauge — tip Korsonic 324 – Cala & Cdria Pipeline Services Company Ltd: Ultrasonic Transducer (UT) is mounted inside the inspection device.

	Axial direction									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	14.20	14.00	13.00	13.00	13.10	13.20	13.60	14.20	14.20	14.20
2	13.20	13.20	13.40	13.40	13.40	13.40	13.60	14.00	14.00	13.40
3	13.60	13.40	13.40	13.80	13.60	13.60	13.80	13.80	13.60	13.20
4	13.60	13.80	13.40	13.60	13.60	13.60	13.40	13.20	13.20	13.00
5	13.60	13.67	13.73	13.80	13.60	13.40	12.80	12.80	12.80	11.60
6	13.50	13.30	12.95	12.60	12.80	13.00	12.80	11.60	11.80	10.80
7	13.40	14.00	13.00	12.80	12.80	12.57	12.33	12.40	11.80	11.60
8	13.60	12.80	13.00	12.80	12.45	12.10	11.87	11.80	11.40	11.00
9	13.80	13.80	13.40	13.20	12.60	12.00	11.40	11.40	11.20	11.00
10	13.80	13.20	14.20	14.00	13.00	12.00	11.20	11.50	11.20	11.00
11	13.60	13.40	13.60	12.80	12.80	12.80	13.40	11.60	11.30	11.00
12	13.40	13.00	12.60	12.10	11.60	11.70	11.80	11.60	11.40	11.10
13	13.40	13.40	13.60	13.80	14.00	13.60	12.80	13.60	12.00	11.50
14	13.60	13.80	13.60	13.40	13.60	13.60	13.80	13.40	12.60	12.10
15	13.60	13.70	13.75	13.80	13.80	13.80	13.60	13.60	13.20	13.40
16	13.60	13.60	13.80	13.80	13.60	13.40	13.40	13.40	13.40	13.40
17	13.60	13.60	13.80	14.20	13.60	13.30	13.50	13.40	13.60	13.20
18	13.60	13.80	13.60	13.60	13.60	13.20	13.60	13.40	13.20	13.00
19	13.40	13.60	13.80	13.50	13.20	13.40	13.40	13.40	13.40	13.40
ACP	13.20	12.80	12.60	12.10	11.60	11.70	11.20	11.40	11.20	10.80
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	12.80	14.20	14.00	13.60	13.40	13.40	14.20	14.20	14.20	14.20
2	13.20	13.00	14.20	13.60	13.20	13.00	13.40	13.00	13.20	13.20
3	13.20	13.20	13.00	13.20	13.40	13.00	13.40	13.60	13.40	13.40
4	13.00	13.00	13.20	13.40	13.40	13.20	13.60	13.40	13.20	13.20
5	11.00	10.80	10.60	11.00	11.40	12.28	13.16	13.50	13.10	13.00
6	10.60	11.00	12.60	11.00	11.30	11.76	12.68	13.60	13.00	13.20
7	11.40	11.20	11.00	11.00	11.20	11.20	11.51	11.83	11.53	11.96
8	11.40	11.00	10.80	10.80	11.00	10.47	10.09	10.06	10.07	10.73
9	11.70	11.70	11.60	11.70	10.56	9.73	9.36	8.98	8.60	9.83
10	11.70	12.40	12.40	11.25	10.12	9.00	8.80	8.60	8.40	9.06
11	11.00	11.00	11.00	10.80	9.88	8.96	8.44	8.32	8.20	8.00
12	10.80	10.60	10.80	10.40	9.60	8.80	8.00	8.67	9.33	10.85
13	11.20	11.40	11.00	11.40	11.07	10.27	9.66	9.80	10.47	11.24

	Axial direction									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
14	11.60	11.60	12.00	12.40	12.53	11.80	11.31	11.27	11.60	12.32
15	13.60	13.60	13.60	13.40	14.00	13.40	12.97	12.73	13.07	13.40
16	13.60	13.40	13.40	14.20	13.40	13.40	13.80	14.20	13.40	13.40
17	14.20	13.20	13.20	13.20	13.60	13.60	13.40	13.60	13.30	13.90
18	13.80	13.40	13.40	13.40	13.20	13.40	13.50	13.00	13.20	14.20
19	13.40	13.60	13.40	13.20	13.60	13.40	13.60	13.60	13.60	13.60
ACP	10.60	10.60	10.60	10.40	9.60	8.80	8.00	8.32	8.20	8.00
	21	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
1	14.20	13.40	13.20	14.20	13.90	13.60	13.60	14.20	14.20	14.20
2	13.20	13.00	13.00	14.20	13.20	13.20	13.70	14.20	14.20	13.60
3	13.20	13.20	13.60	13.20	13.40	13.40	13.40	13.00	12.40	13.40
4	13.40	13.80	13.20	13.20	13.40	13.80	13.30	13.40	13.40	13.40
5	13.00	13.00	12.80	13.60	13.40	13.60	13.20	13.20	13.20	13.60
6	13.00	12.80	13.60	13.00	13.20	13.40	13.20	13.20	13.20	13.20
7	12.85	13.20	14.20	13.20	13.40	13.20	13.40	13.40	14.00	13.40
8	11.62	12.51	13.40	13.40	13.20	13.60	13.20	13.80	13.60	13.40
9	11.05	12.28	13.50	13.50	13.50	14.00	13.60	13.60	13.60	13.60
10	10.57	12.09	13.60	13.60	13.80	13.40	13.40	13.40	13.60	13.20
11	10.85	13.70	13.00	13.20	13.40	13.40	13.60	13.20	13.40	14.20
12	13.80	13.20	13.40	13.40	13.40	13.40	13.60	13.90	14.20	13.40
13	13.50	13.20	13.40	13.40	13.80	13.60	13.60	13.60	13.60	13.40
14	13.20	13.60	13.80	13.80	13.60	13.40	13.60	14.20	13.40	13.30
15	13.60	13.60	13.90	14.20	13.60	13.40	13.47	13.80	13.20	13.20
16	13.40	13.80	13.80	13.60	13.20	13.50	13.47	13.40	13.40	13.90
17	14.20	14.00	13.80	13.40	13.40	13.60	13.60	13.40	13.30	13.55
18	14.20	13.40	13.40	13.40	13.60	13.40	13.30	13.20	13.20	13.20
19	13.60	13.40	13.40	13.80	13.20	13.40	13.20	13.25	13.30	13.35
ACP	10.57	12.09	12.80	13.00	13.20	13.20	13.20	13.00	12.40	13.20

	31	**CCP**	**Defect characteristics**
1	13.40	12.80	ACP axial critical profile
2	13.60	13.00	CCP circumferential critical profile
3	13.20	12.40	Pressure max. 6.4 MPa; Temperature 20°; Internal diameter 473.6 mm
4	13.20	13.00	Nominal thickness 14.2 mm; Uniform loss of material LOSS 0.77 mm
5	13.20	10.60	Corrosion allowance FCA 1.524 mm; API 5L X-52 steel material

	31	CCP	Defect characteristics
6	13.30	10.60	Conventional extension limit: Sc = 360 MPa
7	13.40	11.00	Ultimate tensile strength of thematerial 490–620 MPa
8	13.40	10.06	Percentage elongation after break: A2 in = 22%;
9	13.40	8.60	Modulus of elasticity (Young): E = 205,000 MPa
10	13.50	8.40	Transverse contraction coefficient (Poisson); μ = 0.3
11	13.60	8.00	Safety coefficient Cs = 1.4
12	13.60	8.00	Allowable resistance Sa = Sc/cs = 344.75/1.4 = 257.14 MPa
13	13.60	9.66	Distance to the nearest discontinuity Lmsd = 700 mm
14	13.60	11.27	RSFa accepted allowable resistance coefficient = 0.9
15	13.60	12.73	
16	13.00	13.00	
17	13.10	13.10	
18	13.20	13.00	
19	13.40	13.20	
ACP	13.00		

Table 1. The matrix of the measured thicknesses corresponding to a specific defect (**Figure 2**).

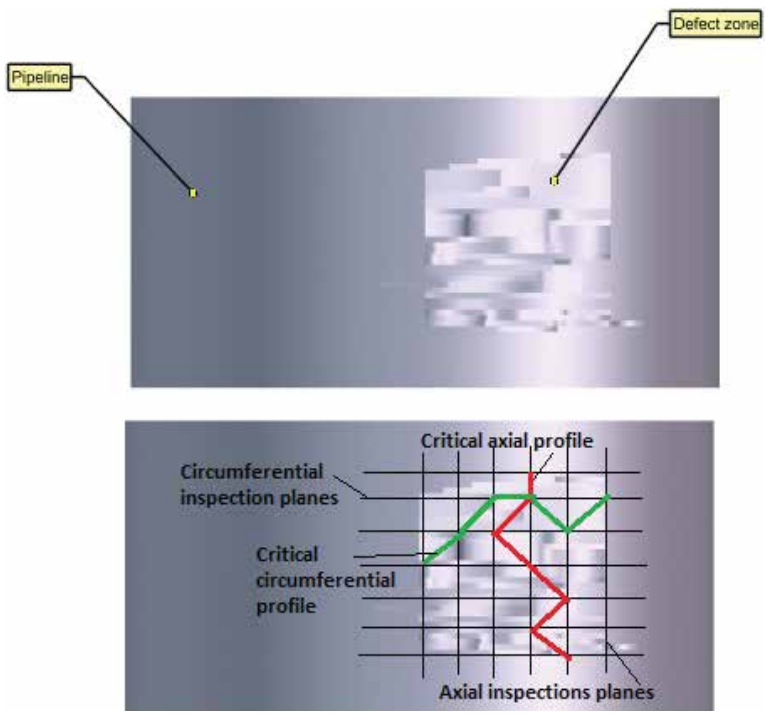


Figure 2. The results of the inspection: tridimensional images of the defects.

(column 1), distance between weld and defect (column 2), the distance from the defect to equipment (column 3), the thickness of the wall of the pipeline at defect position (column 4), clock orientation (column 5), the length of the defect (column 6), the width of the defect (column 7), maximum depth of the defect (column 8), average depth of the defect (column 9), the type of defect (as the cause) (column 10). Where eliminated from the analysis 1662 defects whose causes (column 10) were manufacture, construction activities, repairs, accidental interventions because: They have shallow depths below 20% of the thickness of the pipeline and not due to corrosion, so their development in time is not probable. The remaining 55,162 defects are the following types (**Table 2**): general metal loss, spots, axial and circumferential groove.

As it is seen a large number 55,162 corrosion defects reported by the inspection, but many of them are superficial. Defects have been chosen only to the depth of more than 20% of the thickness the pipeline; their number is 212. The geometrical elements of these defects are in **Table 3**. For the variables *D*, *UTS* and *PO* we considered the values from **Table 4**, [11, 21, 23]. After 8 years of operation ($T_0 = 8$ years), the probability of failure *FP* versus operating pressure is represented in **Figure 3** (calculated at the end of each working year).

The fault location is important to value of the probability of failure FP. If we consider the operating pressure of the pipeline: FP is 0.03 for PO of 5 MPa and equal to 1 for the other pressures of operation of the pipeline. If we consider pipeline pressure at fault position then for PO of 5 and 7 MPa, FP is zero, at 9 MPa FP is equal with 0.24 and equal with 1 at 11, 13 and 15 MPa. If we choose the limit of the probability of failure of 0.5 (highly conservative methods of calculus [4] justifies this value), the first case of assessment tells us that at the work pressures above 5 MPa we could not use the pipeline. The second case of assessment tells us that we can use the pipeline at the pressures of 7 and 9 MPa, too. We have thus a lower degree of conservatism. Based on the considerations we made, it can be appreciated the defect evolution in time. It is true that these considerations include several simplifying assumptions, but also includes the results of PIG measurements. In the situation where it is considered the pipeline

Defect type	**Characteristics**
General loss of metal	$\{[w \geq 3A] \text{ and } [l \geq 3A]\}$
Circumferential notch	$\{[w \geq 1A] \text{ and } [0 < l < A]\}$
Axial groove	$\left\{[1A \leq w < 3A] \text{ and } \left[\frac{l}{w} \geq 2\right]\right\}$
Circumferential groove	$\left\{\left[\frac{l}{w} \leq 0.5A\right] \text{ and } [1A \leq l < 3A]\right\}$
Pin	$\{[0 < w < 1A] \text{ and } [0 < l < 1A]\}$
Axial notch	$\{[0 < w < 1A] \text{ and } [l \geq 1A]\}$
Spots	$\left\{\begin{array}{c}[1A \leq w < 6A] \text{ and } [1A \leq l < 6A] \text{ and } \left[0.5 < \frac{l}{w} < 2\right] \\ \text{and not } ([w \geq 3A] \text{ and } [l \geq 3A])\end{array}\right\}$

The geometric parameter A is defined as follows: if $0 < t < 10$ mm then $A = 10$ mm, if $t \geq 10$ mm, then $A = t$, according to Ref. [2]; l defectlength; w defectwidth.

Table 2. The classification of corrosion defects after geometrical aspect.

Reference distance of welding [m]	Distance of defect to welding [m]	Distance to equipment [m]	Wall thickness [mm]	Clock orientation of defect	Length of defectL [mm]	Width of defectL [mm]	Maximum depth of defect [%]	Average depth of defect [%]	Type of defect
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
19.1	−6.12	2	14.2	8:02	23	18	8	5	Axial groove
28.04	−1.39	1.4	14.2	12:14	11	18	14	10	Axial groove

Table 3. The Excel file filled with data analysis elements (values of ultrasonic inspection).

Variable	Distribution law	Mean value	Standard deviation
Outer Diameter, *D*	Normal	508 [mm]	14 [mm]
Ultimate tensile strength of thematerial, UTS	Normal	517 [MPa]	30 [MPa]
Pressure of operating, PO	Gumbel	Steps 5; 7; 9; 11; 13; 15 [MPa]	0.1 [MPa]

Table 4. The probabilistic variables used in the simulation.

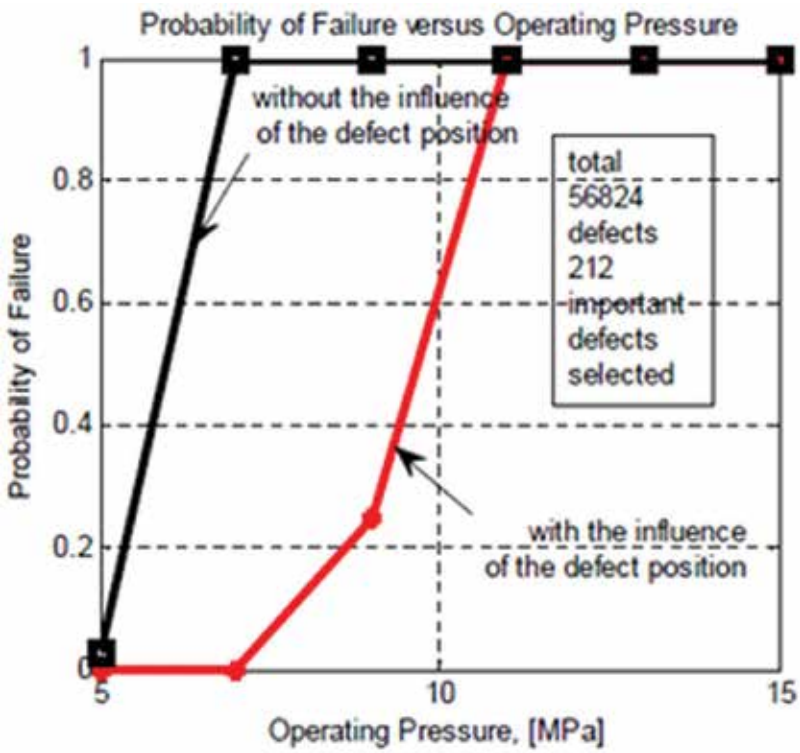


Figure 3. The influence of the defect position versus the probability of failure.

pressure at the defect position was represented the progress of defect probability of failure in time (with a step of 2 years) **Figure 4**. If the pipeline operating pressure is 5 MPa is observed that after 14 years FP grow rapidly, showing that the operator should perform repairs to the system. In the case of operating pressure of 7 MPa since the 12th year of exploitation, FP increases of and between years 12th to 14th FP rises further reaching value 1. Operation of the pipeline to 9 MPa shows a probability of failure which reaches 1 (sure failure) between years from 8th to 12th. Obviously, in situations where the probability of failure is high (it has chosen the 0.5 limit) should intervene to repair the pipeline. By choosing this limit, some defects become critical. The list of defects which should be repaired is given in **Table 5**. If these defects are repaired, then the pipeline FP falls, and it can be used safely for many years, and over a range of operating pressures as shown in **Figure 5**. The method described above was implemented on programs (in Matlab) made by the authors [22, 23]. It is generated a list of defects to be repaired every year, **Table 5** (an example for two pressures 5 and 7 MPa). If the number of the defects is high, an economic analysis of whether a repair or a replacement of the section of the pipeline is required. So we have a procedure of action based on the results of inspection and the accomplished analysis, useful in the maintenance process. The effect of the repair is seen immediately; the pipeline is less likely to fail. However, at higher operating

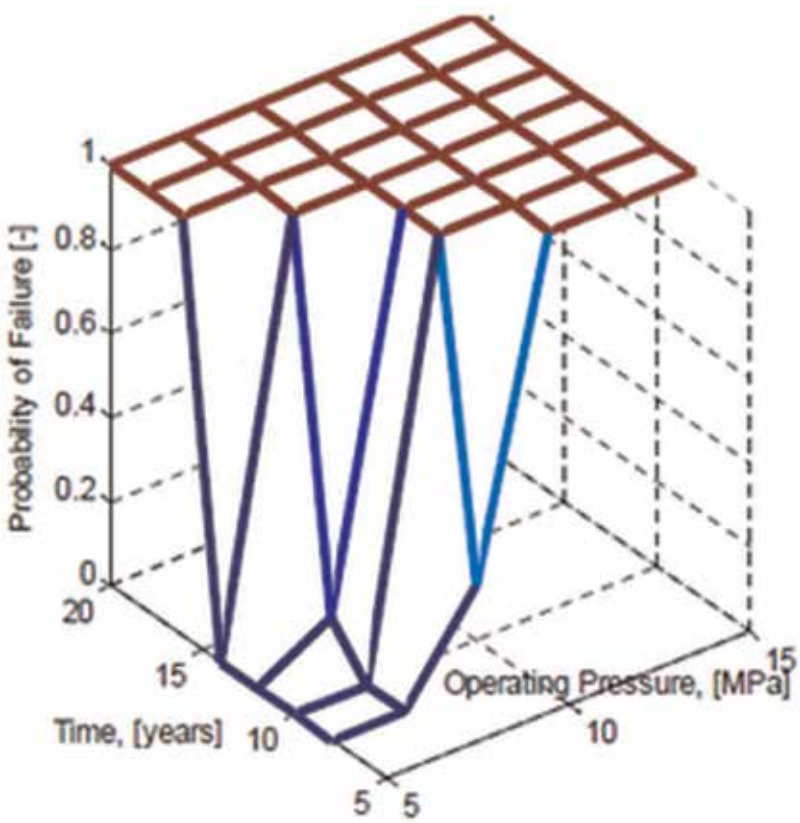


Figure 4. Probability of failure depending on operating pressure and time, if the critical defects are not remedied.

Year	The working pressure 5 [MPa], the number of defects to be repaired 15	The working pressure 7 [MPa], the number of defects to be repaired 33
	The number[1] of the defect repaired	**The number of the defect repaired**
12	–	37 102
14	–	22 25 104 187
16	37 102 104 187	9 16 45 57 61 99 100 127
18	22 25 57	19 50 72 75 111 126 155 166
20	45 50 61 99 100 127 166 205	17 35 51 55 91 101 114 154 162 165 205

[1]The defects are numbered from 1 to 212.

Table 5. List of the defects with the probability of failure greater than 0.5, which must be repaired at the beginning of each year (example of the maintenance program).

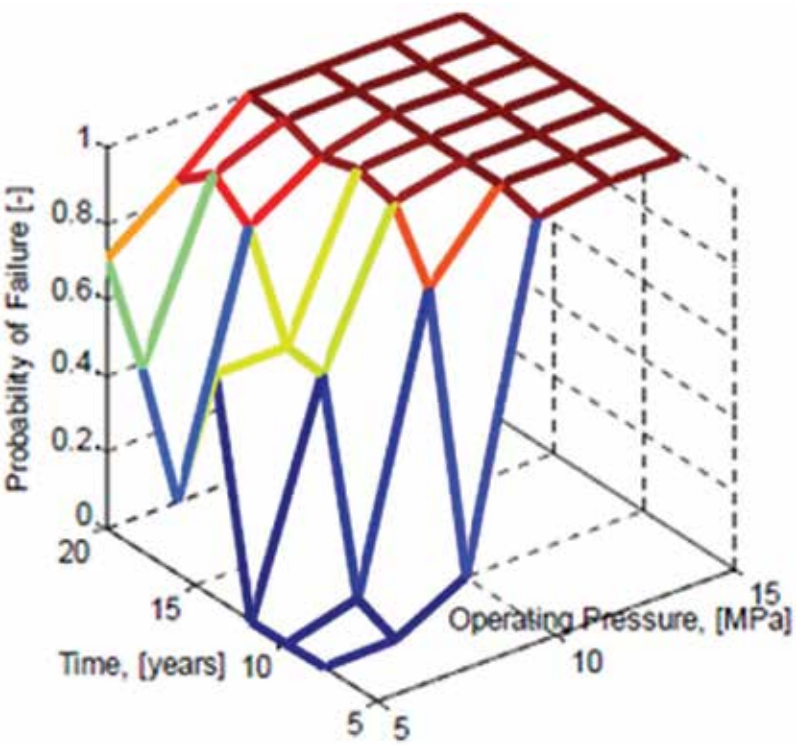


Figure 5. Probability of failure depending on operating pressure and time, when the critical defects are repaired.

pressure, the pipeline conditions lead to a FP equal to 1 regardless of its status. All theoretical models are tested on samples taken from the defective pipeline s. 2.4, to verify the accuracy of the assumptions used [23].

2.4. Experimental determination of pressure burst for the pipes with faults type losses of material

Experimental verification of the behavior of the mechanical elements of pipeline with defects of various kinds is one of the methods for establishing the reserve strength of mechanical resistance. It can draw conclusions about the level of trust that must be attached to the results of assessing the seriousness of defects by the available analytical methods. Studies should include up to burst pressure test of the pipe that have been identified local defects such as loss of material [8]. The following example shows how to perform a test (external diameter $De = 508$ mm, wall thickness $s = 6.3$ mm) of a specimen, which was taken from a section on which were discovered defects of the type material loss. **Figure 6** has revealed defects of the type of material loss that were discovered on that section after inspection with smart devices by type PIG. The geometrical characteristics of defects of the type material loss from the sample under test pressure are given in **Table 6**.

The sample for internal pressure testing consists of a fragment cut from a pipe and two bottoms dished welded ends, which were mounted two connections: the first for ventilation of the sample prior to pressurizing and subsequently for manometer mounting and the second for filling of the

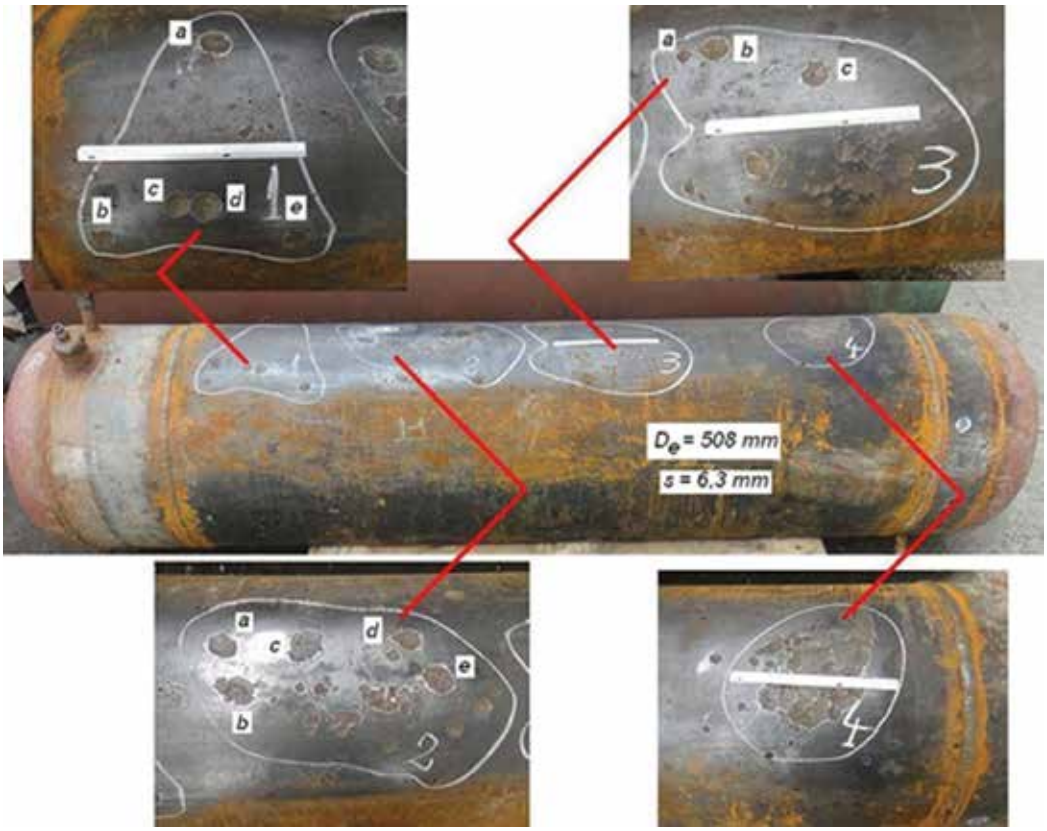


Figure 6. Defects loss of material from the sample being tested.

		Geometric characteristics of the defect		
Location area	**Defect code**	**Maximum depth, d, [mm]**	**Circumferential extension, l_c [mm]**	**Axial extension, l_a [mm]**
1	1a	3.3	35	30
	1b	1.7	25	25
	1c	2.6	25	25
	1d	3.1	30	30
	1e	2.2	25	25
2	2a	3.2	35	35
	2b	5.0	50	35
	2c*	**5.0**	**40**	**45**
	2d	3.6	40	30
	2e	3.9	40	35
3	3a	1.6	15	15
	3b	4.1	35	40
	3c	3.6	30	40
4	**4a****	**4**	**3.2 … 4.8**	**190**

*Defect to which the tenso resistive transducer was glued.
**Zone wherein the breaking of the sample occurred.

Table 6. The geometric characteristics of the defects type loss of material subjected to the burst pressure test.

sample with water and pressurizing it. The stand that has been tested the sample to internal pressure (up to burst) was conducted at the Petroleum—Gas University of Ploiesti and reproduce diagram in **Figure 7**, which presents the constructive elements: the high pressure pump and a platform working in the organization of the stand, **Figure 8**. While conducting the experiment, on the sample of pipeline were applied around the fault with code 2c (defect considered to be the most dangerous depending on the geometrical characteristics), transducers in two directions, circumferential direction TER 1 and the axial direction TER 2. During work, the computer controls data acquisition by using the SPIDER 8 by means of specialized software Catman, which has multiple facilities on determining the number of channels, frequency of data acquisition, storing them in formats that allow the processing with specialized programs, etc. Results of the experimental analysis by resistive tensometry are summarized in the graphs in **Figures 9** and **10**.

Processing of the experimental results was performed as follows: Mechanical tensions were measured in the circumferential direction $\sigma_{\theta ij}$ and axial direction σ_{zij}, using known formulas [24]:

$$\sigma_{\theta ij} = \frac{E}{1-\mu^2}(\varepsilon_{\theta i} + \mu\varepsilon_{zj}) \tag{11}$$

$$\sigma_{zij} = \frac{E}{1-\mu^2}(\varepsilon_{zj} + \mu\varepsilon_{\theta i}) \tag{12}$$

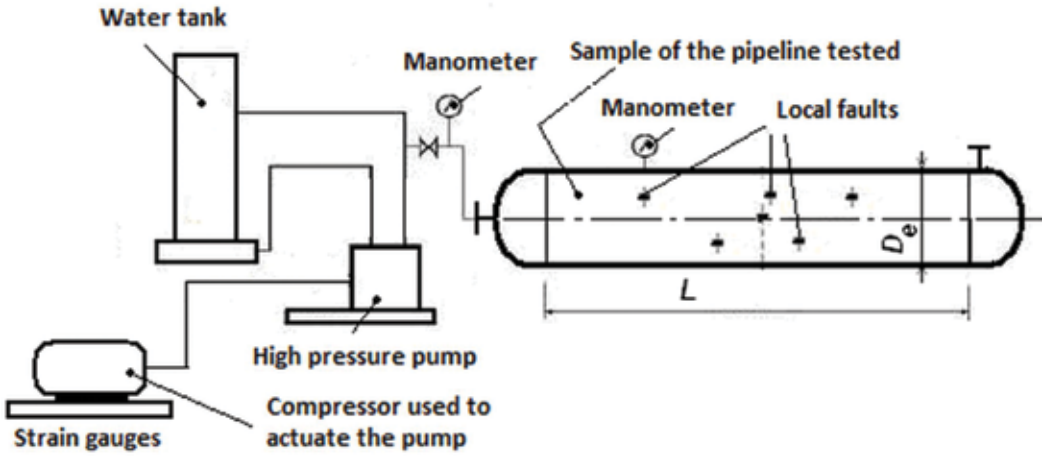


Figure 7. Sketch of the stand used to inner pressure test of pipe samples with local surface defects type loss of material.

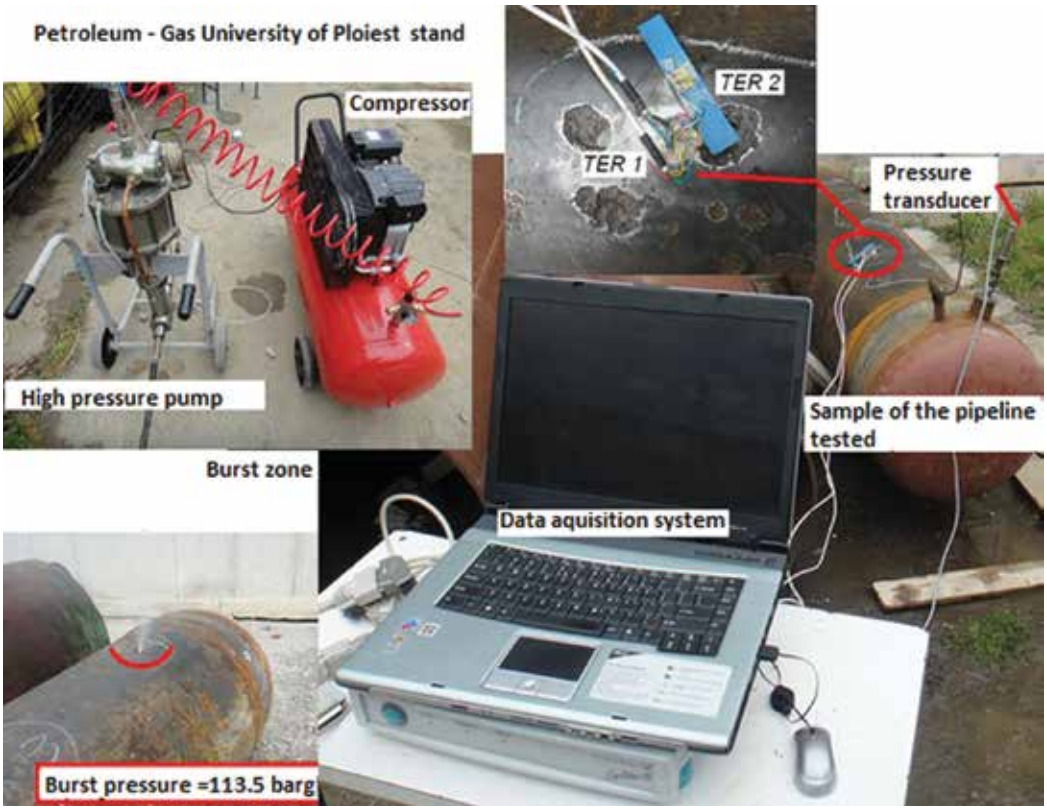


Figure 8. Main components of the stand used to inner pressure test of pipe samples.

http://dx.doi.org/10.5772/intechopen.68227

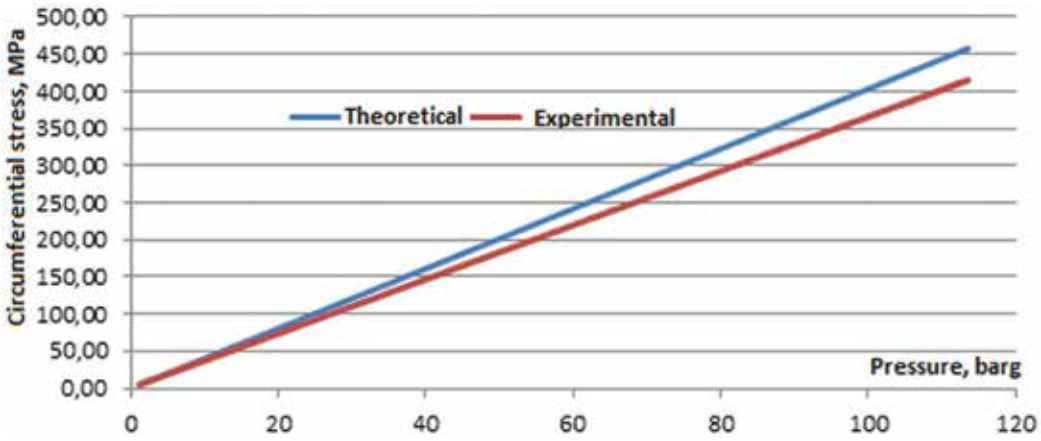


Figure 9. The results of experimental analysis (method of resistive tensometry) for sample with defects type material loss: circumferential stress.

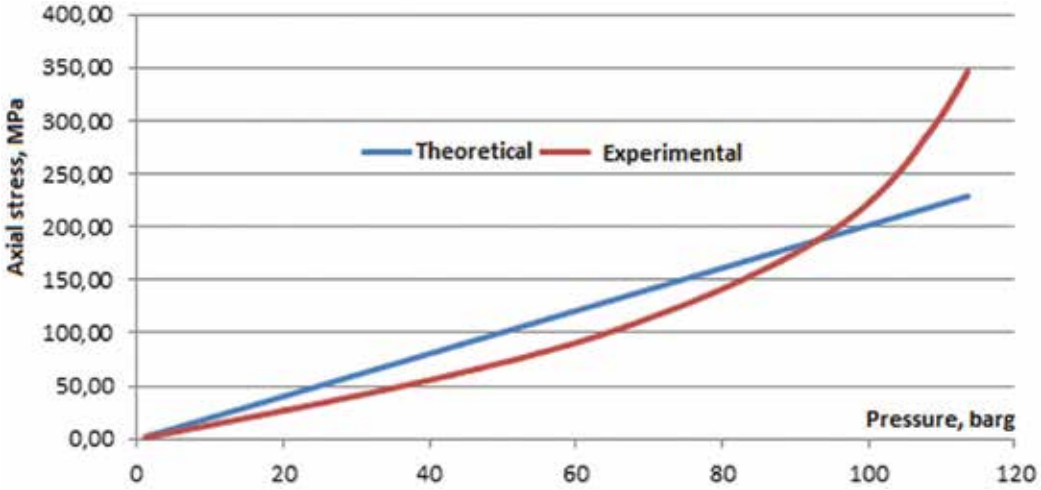


Figure 10. The results of experimental analysis (method of resistive tensometry) for sample with defects type material loss: axial stress.

where E is the longitudinal modulus of elasticity and μ is Poisson's ratio for steel sample; $\varepsilon_{\theta i}$ is specific deformation in the circumferential direction; ε_{zj} is specific deformation in the axial direction (i and j are the identification numbers of transducers). It have been built experimental dependencies of the circumferential and axial deformations shown in **Figure 9**, respectively, **Figure 10**, and these dependencies were compared with the theoretical ones σ_θ, σ_z as described by the formulas:

$$\sigma_\theta = \frac{p_p \cdot D_e}{2t} \tag{13}$$

$$\sigma_z = \frac{p_p \cdot D_e}{4t} \tag{14}$$

In conclusion, experimental verification of the behavior of the mechanical elements of pipeline sets the level of confidence to be associated with the results of assessing the seriousness of defects by analytical methods; the stand designed and built at the Petroleum—Gas University of Ploiesti allows research concerning the pipes behavior with or without defects and can provide results obtained using electro transducers—strain gauges applied to the sample and the sample burst pressure.

3. Conclusions on the case studies

As mentioned in Section 2.1, the case study presented it is focused on practical aspects for the maintenance of pipeline systems. Total length of the pipelines in Romania national gas transport system is 14,500 km and for liquid petroleum products 6000 km. These pipes have a lifetime of between 8 and 45 years, and most of the pipelines have been installed before year 2000. Therefore, the transporters have a real problem with the defects that have appeared over time. Solving them in economic conditions is a difficult problem for these companies. Valorization of PIG inspection results in economic conditions involves selecting of defects. The studies and collaborations of authors with the national companies Transgaz SA and Conpet SA led to implementing maintenance programs (based on the foregoing ideas) that involved the reduction of expenses. Faults evaluation was based on the pressure at the defect position, which reduced the degree of conservatism and maintenance costs. We can do that because we know the position of the defect. The authors of this paper are also working with NRAE from Romania organization dealing with energy issues in Romania. We propose to be added to the indicators that relate to the safety of hydrocarbon transport systems, the probability of failure (with relation from 2.2, an example is introduced in 2.3). This indicator compared to the number recorded accidents through inspections, the number of accidents reported by third parties, the number of accidents that occurred on km of pipe, is a prediction, helping to increase security in transport systems. A procedure was made and submitted for discussion. The role of the experiment in certifying the results and building the trust of the beneficiaries (TSOs) s. 2.4 is underlined by describing the stand used for testing samples of pipeline sections.

Currently, technique for determining the image of defects in the pipelines achieved results increasingly better [25, 26]. On the basis of tridimensional images of the defects can make the three types of analysis described in the introductory part.

The analyzes use only defect length and depth of the defect (level 1); critical profile of the defect in the longitudinal and transverse direction (level 2) **Figure 11a** and **b**; a tridimensional image of the defect generated based on readings during inspection stage, transformed into a solid object and subjected to finite element analysis (level 3) **Figure 11c–e**. Based on the defect report (**Table 1**), we can achieve a three-dimensional model, using a mesh (**Figure 11c**) and the analyze with finite element method (**Figure 11d**). In **Figure 11**, **e** is a detail with defect zone.

The technique described in this case study uses also other features representing the variation of parameters of evaluation: material characteristics; the initial thickness of the pipe; operating pressure. The main objective is to intervene in economic conditions to repair a defect that evolves over time.

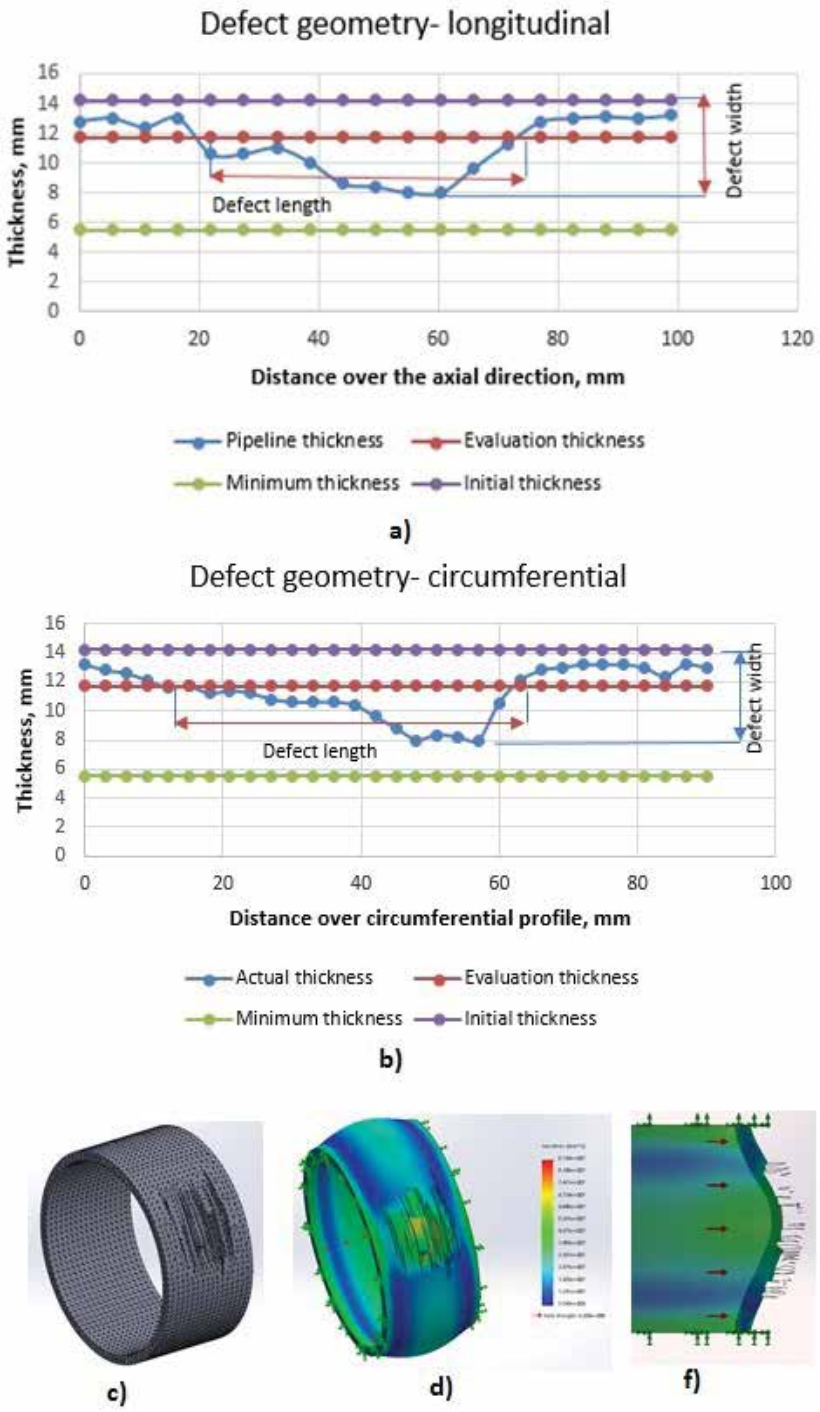


Figure 11. Using different levels of analysis: (a) critical axial profile used in a Level 2 analysis; (b) critical circumferential profile used in a Level 2 analysis; (c) treedimensioal model and the mesh of finite elements for a Level 3 analysis; (d) simulation model (solidworks simulation); a detail with the defect zone.

The truth is that we do not know very well the moment. There are many influences including corrosion rate, an important parameter with a variation difficult to estimate. If we compare with a real-life situation in which political decisions (which are based in many cases on less knowledge) can affect the lives of millions of people, we are still much better.

We know the shape and location of the fault. Finite element analysis seems to be the best method of assessment and perhaps soon a probabilistic assessment, showing that the three-dimensional shape change of the defect in time will be possible. Therefore, the precision regarding of best time for intervention into the system, to avoid a critical situation, it will be increased.

In conclusion, we consider that the main contribution of the article is to transform ideas, indications of standards and inspection data in a coherent system to prevent critical situations, in economic conditions.

Author details

Mihail Minescu and Ion Pana*

*Address all correspondence to: ion.pana@upg-ploiesti.ro

Petroleum and Gas University from Ploiesti, Faculty of Mechanical and Electrical Engineering, Romania

References

[1] Dinu F. Extraction and Processing of Natural Gas. 2nd ed. Romania, Ploiesti: Petroleum and Gas University Publishing House; 2013. p. 352

[2] API 579-1/ASME FFS-1 2007 Fitness-For-Service. Washington, D.C: American Petroleum Institute; 2007.

[3] API 579-1/ASME FFS-1 2009 Example Problem Manual. Washington, D.C.: American Petroleum Institute; 2009.

[4] B31G / ASME 2012 Manual for Determining the Remaining Strength of Corroded Pipelines. USA: American Society of Mechanical Engineers; 2012. p. 60

[5] Shekari E, Khan F, Ahmed S. A predictive approach to fitness-for-service assessment of pitting corrosion. InternationalJournal of Pressure Vessels and Piping. 2016;**137**:13–21

[6] Scano L. Fitness-for-service assessment of a steam pipeline operating in the creep range after 230000 hours of service. In: Proceedings of ASME 2012 Pressure Vessels and Piping Conference, Vol 3: Design and Analysis; 15–19 July; Toronto, Ontario, Canada. 2012. pp. 626–632

[7] Almeida JIL, Rodrigues MC, Mishina KDV. Modeling and level 3 fitness-for-service assessment of a cylindrical pressure vessel with general metal loss in conjunction with the numerical thermal transient analysis. InternationalJournal of Mechanics and Applications. 2014;**4**:80–93. DOI: 10.5923/j.mechanics.20140403.02

[8] Janelle JL. An overview and validation of the fitness-for-service assessment procedures for local thin areas[thesis].Akron, Ohio, United States: University of Akron; December, 2005.

[9] Adib-Ramezani H, Jeong J, Pluvinage G. Structural integrity evaluation of X52 gas pipes subjected to external corrosion defects using the SINTAP procedure. International Journal of Pressure Vessels and Piping. 2006;**83**:420–432

[10] Ahammed M. Prediction of remaining strength of corroded pressurized pipelines. InternationalJournal of Pressure Vessels and Piping. 1997;**71**:213–217

[11] Ahammed M. Prediction of remaining strength of corroded pressurized pipelines. International Journal of Pressure Vessels and Piping. 1998;**75**:321–329

[12] Li SX, Yu SR, Zeng HL, Liang R. Predicting corrosion remaining life of the underground pipelines with mechanically based probabilistic model. Journal of Petroleum Science and Engineering. 2009;**65**:162–166

[13] Netto TA, Ferraz US, Estefen SF. The effect of corrosion defects on the burst pressure of pipelines. Journal of Constructional Steel Research. 2005;**61**:1185–1204

[14] Teixeira AP, Soaresa CG, Nettob TA, Estefen SF. Reliability of pipelines with corrosion defects. International Journal of Pressure Vessels and Piping. 2008;**85**:228–237

[15] Minescu M, Pana I. The equivalence of the assessments procedures API 579 and ASME B31G. Petroleum and Gas Bulletin Technical Series. 2012;**64**:11–20

[16] DNV-RP-G101 2010 Recommended Practice Risk Based Inspection of Offshore Topsides Static Mechanical Equipment. Det Norske Veritas. October 2010. Available from: http://www.dnv.com [Accessed: 2016]

[17] Dinu F. Extraction and Treating of Natural Gas, 1st ed. Romania, Ploiesti: Petroleum and Gas University Publishing House; 2009. p. 256

[18] Caleyo F, Gonzales JL, Hallen JM. A study of the reliability assessment methodology for pipelines with active corrosion defects. International Journal of Pressure Vessels and Piping. 2002;**79**:77–86

[19] De Leon D, Macias OF. Effect of spatial correlation on the failure probability of pipelines under corrosion. International Journal of Pressure Vessels and Piping. 2005;**82**:123–128

[20] Mustaffa Z, Gelder P. A review and probabilistic analysis of limit state functions of corroded pipelines. In: Proceedings of the Twentieth International Offshore and Polar Engineering Conference; 20–25 June; China, Beijing. 2010. pp. 626–632

[21] Zecheru G, Neacsu M, Lata I. The probabilistic conception of the pipeline defect acceptance charts. Fracture Mechanics Bulletin ARMR. 2005;**17**:31–36

[22] Minescu M, Pana I. Safety evaluation of the pipelines systems for petroleum products. Applied Mechanics and Materials. 2013;**371**:782–786. DOI:10.4028/www.scientific.net/AMM.371.782

[23] Minescu M, Pana I, Dumitrescu A. Implementation of an Assessment Procedure Type Fitness-For-Service in Accordance with API Standard 579, for Oil Pipelines Faults and Technical Assessment of a Real Pipeline Section [research service contract no. 387/09.12.2010, beneficiary S.C. Conpet SA Romania], Petroleum & Gas University Ploiesti; 2011

[24] Stanciu LS, Popa I. Practical Works and Calculation for Mechanical Vibration. 1st ed. Romania, Ploiesti: Petroleum Gas University Publishing House; 2015. p. 200

[25] Barbian A, Beller M, Hugger A, Jäger C, Pfanger A. Threshold, accuracies, and Resolution, Proceedings of the Pipeline Pigging and integrity Conference; 15–18 February; Houston, Clarion, 2010

[26] Barbian A, Beller M, Hartmann S., Schneider U. High resolution ultrasonic in-line inspection: Added value and special applications. 6th Pipeline Technology Conference; 4 April; Hannover. Germany, Hannover Messe, 2011

Chapter 13

Automated Fault Detection System for Wind Farms

Deepshikha Agarwal

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.68844

Abstract

Due to the pollution and health hazards of nonrenewable resource-based energy generation systems, now focus is on the use of renewable resources. This chapter aims as providing an automated fault-detection system for increasing the robustness of offshore located wind farms. The method is based on the use of flexible threshold for calculation of the collected sample values. A fuzzy inference system (FIS) is designed for the automatic real-time fault detection system named as FIS-based fault detection system (FFDS) for offshore wind farms. The method uses the concept of combination-summation (CS) and flow-directions to determine the extent of fault occurrence in the wind farm. Based on the working conditions of the wind farm, preventive or corrective measures are suggested to the remote observer. The performance of these methods is evaluated on MATLAB.

Keywords: wind farm, wireless sensor networks, threshold, fault detection, network lifetime

1. Introduction

Wind energy is freely available everywhere in abundance. It is a renewable resource that will never get exhausted. This energy if properly utilized can lead to greener and safer energy generation compared to coal generated electricity. It is also one of the lowest priced renewable energy technologies available nowadays [1].

In 2015, energy produced in the United States was about 91% of U.S. energy consumption due to less import of petroleum [2]. Majority of energy production being due to fossil fuels, i.e., coal, petroleum, and natural gas. According to Ref. [2], natural gas contributed 32% of total generation, petroleum 28%, coal 21%, renewable energy 11%, and nuclear electric power 9%.

However, using natural gas for energy generation has several issues. First, leakage of methane during drilling and extraction of natural gas from wells and its transportation in pipelines [3]. Methane is stronger than CO_2 at trapping heat and causing global warming. Methane emissions range from 1 to 9% of total life cycle emissions. Natural gas-fired power plants contribute to acid rain and ground-level ozone, both of which can damage forests and agricultural crops [4].

The present renewable energy-based generation plants such as offshore wind farms are not entirely capable of fulfilling the future needs of the society. Due to this reason, wind-based energy generation is still not very popular and is unable to replace coal or natural gas-based energy production. The monitoring and control systems used are now obsolete and new methods are required.

The control and maintenance actions require complete human interference, and it is a time-consuming process. These challenges lead to extra cost on emergency maintenance, component screening, and physical designs.

Wind turbines consist of several components and are subject to various failures of electrical and mechanical nature [5], e.g., imbalance in electrical controls, gearbox, and yaw system. Some are more frequent and cause larger downtime of the whole system. These faults cause rotor imbalance, unbalances and harmonics in air gap flux, increase torque pulsation, and increase losses and reduction in efficiency by directly affecting the power, current, and voltage output of the generator. Therefore, monitoring of these critical components should be on the highest priority so that plant downtime can be reduced. The offshore located wind turbine generator system requires monitoring of parameters such as sea-surface temperature, wind velocity, water salinity, wave heights, and strain measurement [6, 7]. However, the monitoring of wind turbine parts has several practical difficulties, e.g., limited accessibility, large size and complex geometry of the blades, effect of environmental parameters, etc.

Several papers have discussed methods to detect faults in wind farms, e.g., gearbox fault detection using discrete wavelet transformation [8]. Similarly, high frequency vibration data collected from gearbox testing were used to gearbox fault detection in Ref. [9], which included k-means clustering algorithm. The drawbacks of this system are the assumption that the underlying process is stationary and the time factor is eliminated. Brandão et al. [10] discuss neural networks for fault forecasting of wind turbine gearbox. Badihi [11] presents protection of against the decreased power generation caused by turbine blade erosion and debris on the blades. A fault diagnosis method based on signal analysis and recognition is presented [12]. Time-frequency representations have been proposed in the literature [13–15]. These techniques have high complexity and poor resolution [16]. One approach used Hilbert transformation in a doubly fed induction generator-based wind turbine [17].

Hence, there should be some automated systems to remotely monitor these parameters and notify about faults in the system. By using wireless sensor networks (WSNs), we can ensure reliable operation of wind farm. This helps in reducing manual interference and wind farm can be completely monitored for 24 hours every day. The following sections discuss how this can be performed.

http://dx.doi.org/10.5772/intechopen.68844

2. Flexible threshold selection scheme

In the past, the monitoring systems used constant threshold to record the data independent of time of the day or month. The constant threshold is calculated as the average of the dataset. As a result of several observations, it can be concluded that such a scheme does not give accurate results if there are changes in the scene or environment pertaining to parameters under consideration, e.g., the temperature of air during daytime is higher compared to night time. Similar variation is true during different seasons, e.g., average temperature during winter season is different from the average temperature during summer season. Hence, if constant or fixed threshold value is chosen for the entire dataset, it is likely to give unoptimized results for both the scenes. Moreover, if the chosen fixed threshold value is very high, it will result in many missed detections, and if it is very low, it will lead to many false positives.

Hence, threshold value should be selected using an appropriate scheme that allows dynamic change in the threshold value to accommodate the variations in time of data recording. This method gives better performance in terms of sensed parameters. The threshold provides a reference for finding values that are higher or lower than the threshold both of which may indicate health failures in the wind farm.

The WSN topology in wind farm consists of tower fixed nodes [18]. These are wireless sensor nodes attached to the tower nodes that can continuously sense the parameter values (samples) throughout the day and night. This information is converted into data packets that are transmitted to the sink node by taking multiple hops through the scattered sensor nodes. The sink node is located at the end of the wind farm. Every tower-fixed node is allocated a fixed local unique address called as RTN id (row-tower-node), which is transmitted as an identification of the originator of packet.

Suppose X_D is a set of samples collected by the tower-fixed sensor nodes during the day period, where

$$X_D = \{X_1, X_2, X_3, \ldots, X_i, \ldots, X_N\} \tag{1}$$

and Y_N is a set of samples collected by the tower-fixed sensor nodes during the night period, where

$$Y_N = \{Y_1, Y_2, Y_3, \ldots, Y_i, \ldots, Y_N\} \tag{2}$$

The samples collected during the night period.

The decision of choosing a new threshold for the dataset depends on the correlation between the datasets. The correlation is the measure of the similarity content between the two datasets. If the correlation of the two datasets is high, it means that the two datasets correspond to the similar time duration of the collected data and hence eliminate the need for calculating another threshold for the new dataset. Similarly, low correlation is indicative of large variations and necessitates the calculation of new thresholds for better data interpretation.

The correlation between the two datasets $R(X_D, Y_N)$ can be expressed as [19]:

$$R(X_D, Y_N) = \frac{\sum_{i=1}^{N}(X_i - X_m) * (Y_i - Y_m)}{\sqrt{\sum_{i=1}^{N}(X_i - X_m)^2 * \sum_{i=1}^{N}(Y_i - Y_m)^2}} \tag{3}$$

where X_i and Y_i are the values of datasets X_D and Y_N at "i" time instant. X_m and Y_m are the average values of the datasets, X_D, Y_N, and N is the number of samples in each dataset which should be the same for X_D and Y_N.

Figure 1 shows the scatter plot for wind speed dataset and its computed correlation coefficient. **Table 1** shows the degree of similarity between the datasets depending on the calculated correlation coefficients.

To calculate the thresholds, T_X and T_Y, the method prefers geometric mean of the datasets with "N" samples, instead of arithmetic mean given as below:

$$T_X = \sqrt[N]{x_1 * x_2 * .... * x_N} \tag{4}$$

$$T_Y = \sqrt[N]{y_1 * y_2 * .... * y_N} \tag{5}$$

We consider geometric mean because the datasets are characterized by a majority of similar

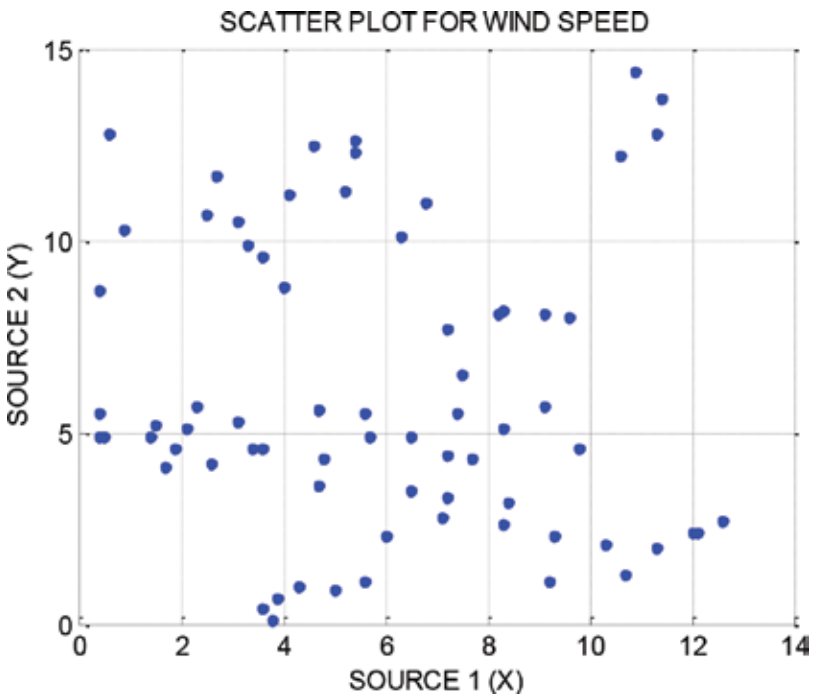


Figure 1. Scatter plots wind speed. Corr_coeff = −0.07846169 [18].

Sl. No.	Corr_coeff.	Degree of similarity	New threshold required
1	−0.7 to +0.7	Low	Yes
2	Less than −0.71	High	No
3	More than +0.71	High	No

Table 1. Criterion for new threshold selection [18].

values and relatively very few extreme values. Being strongly influenced by few extreme observations, the threshold values calculation based on arithmetic means would fail to provide a real means of identifying the extreme values.

The use of a geometric mean normalizes the range being averaged, so that no range dominates the weighting, and a given percentage change in any of the properties has the same effect on the geometric mean. **Table 2** shows the calculated threshold values for the flexible threshold method and the mean method (MM).

Furthermore, the method requires ranging of the infinite sample values into discrete levels without changing the meaning of information using quantization. To do this, first, the distance matrices d_X and d_Y are calculated as below

$$d_X = [X_D - T_X] \tag{6}$$

$$d_Y = [Y_N - T_Y] \tag{7}$$

where these matrices represent the values of X_D and Y_N after thresholding where,

$$d_X = \{x_1', x_2', x_3', \ldots, x_i', \ldots, x_N'\} \tag{8}$$

$$d_Y = \{y_1', y_2', y_3', \ldots, y_i', \ldots, y_N'\} \tag{9}$$

Finally, the quantization is performed on the above values independently with respect to their maximum and minimum values. This can be expressed as

$$\mathrm{Q}\alpha(\max(\mathrm{d}) - \min(\mathrm{d})) \tag{10}$$

where Q is the number of quantization levels for distance matrix d.

SRC	Variation range	Wind speed				
		Range values	Corr. coeff	FTS (T_{FTS})	MM (T_{MM})	$T_{FTS} - T_{MM}$
1(X)	(Min)	0.4000	−0.0784	10.6938	5.9778	4.7160
	(Max)	12.6000				
2(Y)	(Min)	0.1000		12.6125		6.6347
	(Max)	14.4000				

Table 2. Threshold values [18].

The present scenario considers five quantization levels 0, 1, 2, 3, and 4 calculated as from Eq. (10). If the variation in the datasets and the total number of samples in it is large, the number of levels may also increase for better accuracy in fault prediction. This would lead to increase in the size of transmitted packets because the number of bits required to encode each level into binary will also increase. This would cause greater energy depletion in packet transmission, reception by the sensor nodes thus lowering the WSN network lifetime. Thus, the choice of the number of quantization levels should be able to provide accurate fault prediction without compromising the network lifetime. Each level carries a significant and distinct meaning regarding the sensed value, e.g., level "0" indicates that there is no difference between the sample value and the threshold. Similarly, levels "2," "3," and "4" indicate increased levels of variation. **Figure 2** depicts the above method.

2.1. Simulation results and discussion

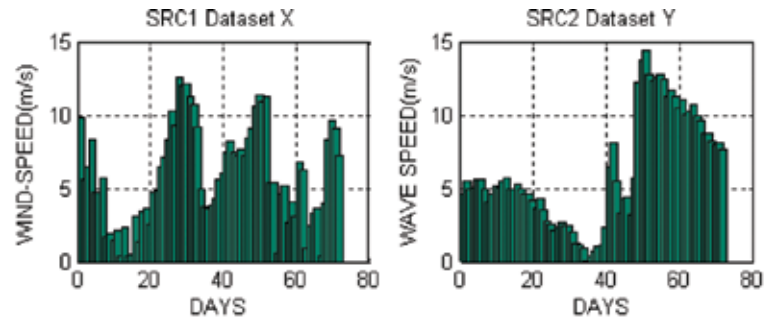
The flexible threshold selection (FTS) method is compared with the mean method (MM). We have considered a total of 72 samples collected during daytime and nighttime for wind speed. The sampling frequency is 1 sample per 10 minutes over a period of 12 hours daytime and 12 hours nighttime. As observed by the simulation results, the flexible threshold method gives a better performance and accurate results for parameter monitoring. **Table 2** depicts the range of collected samples and their calculated thresholds using flexible threshold selection (FTS) and mean method (MM). As observed, the datasets for source 1(X) and source 2(Y) have small variations. If both the datasets from the sources are instead, considered to be one single dataset, the variation of values is large. This causes the static threshold selected using the MM method tends to be biased toward the higher values. However, this is not the case with dynamic threshold. We can calculate different thresholds for datasets collected at different times, which will adapt with the true variations of the values known to nature. Hence, the flexible method is unbiased toward any extreme values and gives a balanced view of the data under consideration. The MM method does not consider computing new threshold every time but it remains unchanged for any dataset making it an unrealistic choice.

Two different thresholds for both the sources find the correlation between them by considering them individually.

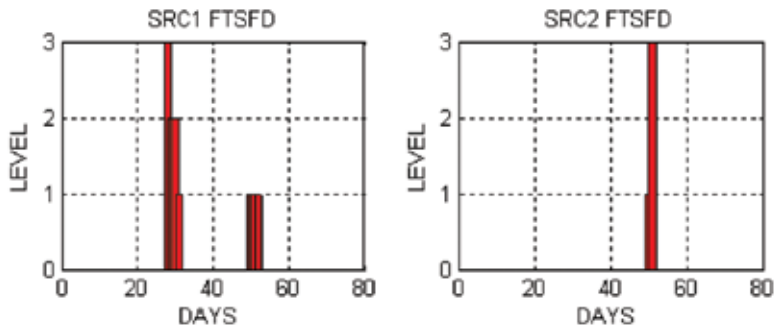
It is clear from the above discussion that the choice of appropriate threshold has a large impact on the quantization levels. The MM method for threshold selection is only able to detect large variation in the values, i.e., levels "2" and "3" whereas in the FTS method the detected levels have a distributed pattern, i.e., it can detect both small and large variations. Also, the levels detected by the FTS method is consistent compared to the MM method, which provides a very accurate status of the conditions of the wind farm.

Figure 2 shows that the FTS method suggests a majority of level "0" occurrences over other levels unlike the MM method where the majority is level "2" occurrences. Thus, it can be concluded that the flexible method is unbiased toward the larger values in the datasets and

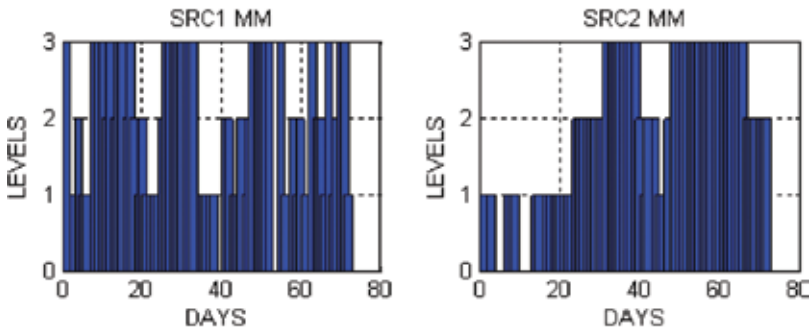
http://dx.doi.org/10.5772/intechopen.68844



(i) Maximum-minimum variations



(ii) FTS method



(iii) MM method

Figure 2. Comparison between MM and flexible method for monitored parameter wind speed [19, 20]. (i) Collected samples of maximum and minimum wave heights, (ii) quantized levels using the FTS method, and (iii) quantized levels using the MM method.

hence provides better accuracy of monitored parameters. The graph in **Figure 2** is generated from real-time data from the Burbon-Nysted wind farm, Denmark.

3. Fault detection scheme

This system is called a fuzzy inference system (FIS)-based fault detection method (FFDS). This is an automated system, gives precise information of the health condition of the wind farm to the remote observer, and gives alarms for taking corrective or preventive measures for maintaining the reliability of the farm.

The observer needs not observe all the properties of the parameter values as a single signal, rather, the degree of similarity between the values finds the basis for choosing a new threshold. This is a simple method that helps in finding real-time data for monitoring purposes. These data when analyzed can predict all possible fault occurrences.

3.1. Automatic fault diagnosis method

The fault detection scheme uses combination-summation (CS) and flow directions (FDs) to design the FIS [9]. This aids to derive significant information from the quantized levels about fault event occurrence in the monitored data samples of offshore wind farm. The received quantized levels corresponding to monitored data samples represent the surrounding environmental conditions in offshore wind farm. The received values being fuzzy in form use FIS to provide accurate interpretation of the environmental conditions.

For determining the CS and FD, five consecutive received levels are considered in one period of time "T" where

$$T = t_1 + t_2 + t_3 + t_4 + t_5 \tag{11}$$

which represents five consecutive time intervals. Depending on the permutation and the summation of the levels, fuzzy logic is used to predict fault occurrences. For example, consider the levels received at "t" times are l_{t_1},l_{t_2}, l_{t_3}, l_{t_4}, and l_{t5} then the summation of levels is

$$\mathrm{CS} = \sum_{i=t_1}^{i=t_5} l_i = l_{t_1} + l_{t_2} + l_{t_3} + l_{t_4} + l_{t5} \tag{12}$$

where the range of CS is [0–20]. The numeral 20 indicates constant occurrence of level 4, i.e., 44,444. The obtained levels can be either repeating or nonrepeating, e.g., 22,222, 31,224, 01,234, and 43,210 as depicted in **Figure 3**. The CS for these levels is 10, 12, 10, and 10. As observed, this alone is not sufficient for fault prediction. Fault prediction can give accurate results if the corresponding FD is also considered with the values. Here, FD means whether the received levels are in state of increasing, decreasing, remaining stable, or varying constantly. For example, if the CS is 10, it has multiple values, but if FD has raising edge, it means the combination suggests fault event occurrence in the future and calls for immediate preventive action. If the levels increase constantly, then FD is considered to be raising edge shown by arrow in upward direction (**Figure 3**).

The remote observer is able to predict meaningful information from these received quantized levels based on the fuzzy-logic rules as presented in **Table 3**.

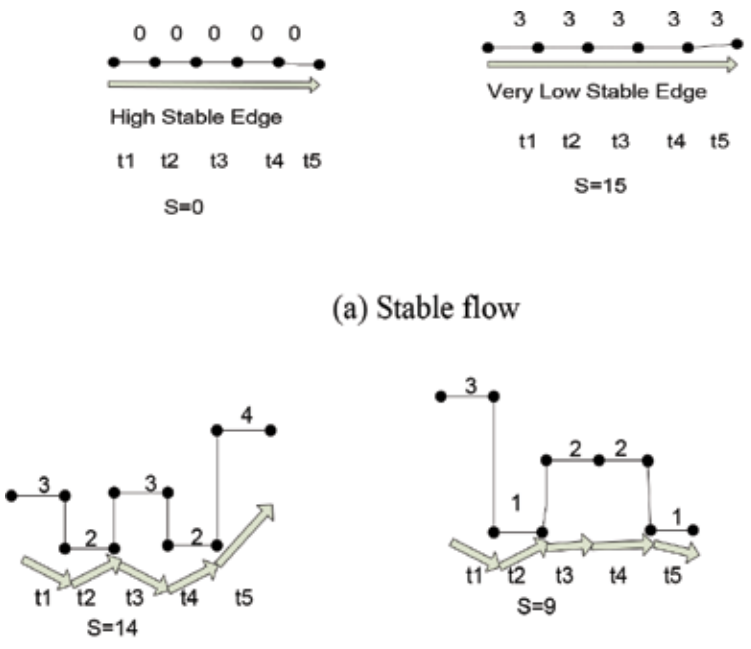


Figure 3. Illustrations of repeating-level and nonrepeating-level combinations. (a) Stable flow, (b) average flow, and (c) raising/dropping flow.

In this system, the fuzzy-set "F" can be described [7, 8] as

$$F = \{\omega, m(\omega) | \omega \in U\} \tag{13}$$

$$U = \{0 - 3, 1\} \tag{14}$$

$$m : \omega \rightarrow [\text{Normal operation, Low risk, High risk}] \tag{15}$$

Sl. No.	Flow directions	Combination-summations	Result (MF)
1	Stable (S)	0, 5	Normal
2	Stable	10	Low risk
3	Stable	15	High risk
4	Stable	20	Very high risk
5	Rising (R)	10	High risk
6	Dropping (D)	10	Low risk
7	Average (A)	1–4	Normal
8	Average	5–6	Low risk
9	Average	7–9	High risk
10	Average	≥ 10	Very high risk

Table 3. Fuzzy rule base.

where ω is the combination-summation and flow-directions of the received levels by the remote observer, $m(\omega)$ is the membership function for the received level and U is the universal set representing the set of all levels, as shown in **Table 3**. The membership function alerts the remote observer whenever the probability of fault occurrence becomes high (**Figure 4**). It can be formulated with risk R_i as

$$(\mathrm{FD_i}, \mathrm{CS_i}) \rightarrow R_i \tag{16}$$

This method is very simple to implement and efficient in enhancing the WSN lifetime.

Table 4 shows the comparison for wind speed data computed from **Figure 2**. The results confirm the belief that FFDS is able to predict accurate conditions of the wind farm. As shown, it predicts normal operation of the farm, whereas MM is only able to detect extreme values of level "0" and "3." This leads to false alarm for corrective measures due to inaccurate calculations. Thus, it can be concluded that the FTS method is unbiased toward the larger values in the datasets and hence provides better accuracy of monitored parameters.

3.2. Simulations and discussion

Table 5 provides the details of simulation parameters used in the study. The sink node is located at the farthest point in the field. **Figure 5** shows the round in which all the nodes in the area become dead (network-lifetime).

The method FTS gives an accurate view of the parameter values in real time and the threshold selection does not indicate any biasing toward a particular value, which is confirmed from **Table 4**. Also, as observed from **Figure 5**, the network lifetime of WSN network is also increased by nearly 10 times with a packet size of 23 bits.

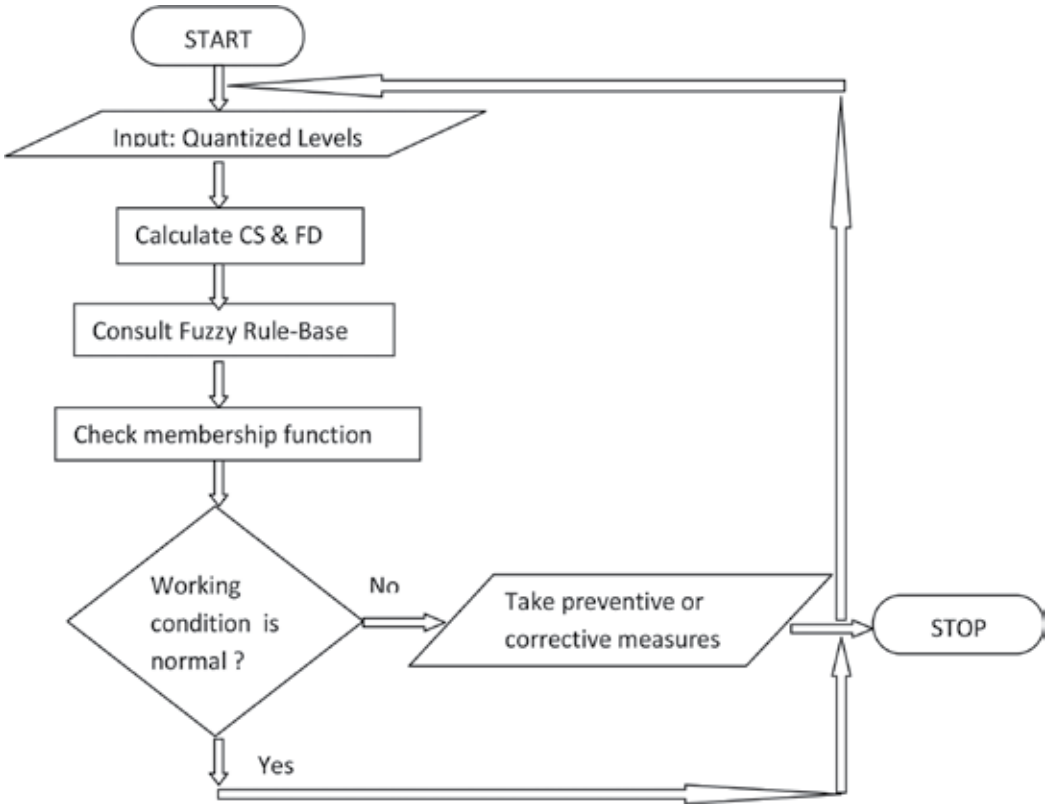


Figure 4. Flowchart to depict working of fuzzy inference system.

Levels	Number of quantized levels		Number of consistent quantized levels			
	FTS	**MM**	**FTS**	**prediction**	**MM**	**Prediction**
"0"	136	19	32	Normal operation	1	Fault has occurred (corrective measures)
"1"	5	37	0		3	
"2"	2	37	0		3	
"3"	1	51	0		9	

Table 4. Wind speed [19].

Sl. No.	Parameter	Value
1.	Area size	1000 m × 1000 m, 2000 m × 2500 m, 300 m × 4500 m, 4000 m × 6000 m
2.	Total number of nodes	416–1216
3.	Total number of fixed nodes	216
4.	Total number scattered nodes	200–1000
5.	Total number of turbines	72

Table 5. Simulation parameters.

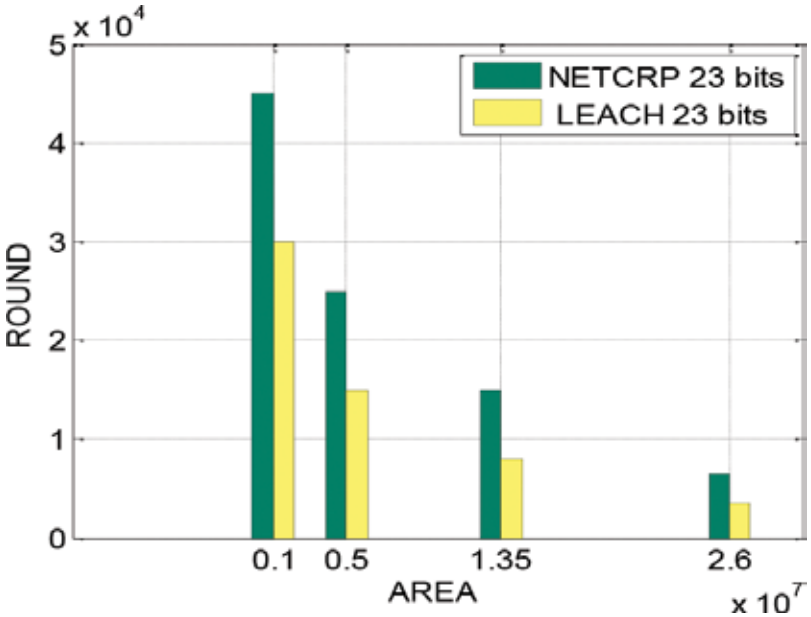


Figure 5. Round in which all the nodes become dead [21].

These observations conclude that the flexible threshold selection method improves WSN network-lifetime by increasing energy savings with respect to earlier methods. Moreover, it is suitable for automated monitoring for all area sizes, large number of nodes and if amount of information to be transmitted is large.

4. Conclusions

This chapter discusses the flexible threshold selection method for efficient environment monitoring of the offshore wind farm. It uses degree of similarity between the previous and the current datasets for calculating geometric mean-based flexible threshold as it does not get biased due to extreme values in the datasets. The method is compared with the static threshold mean method of threshold selection and the performance is seen to be enhanced.

Also, the automated fault detection method is presented in this chapter. This is a simple method that uses small integer values for indicating faulty conditions of the wind farm in real time. The method uses fuzzy inference system that takes integer values as input and gives output in the form of fault status prediction of the farm. Based on these predictions, the system suggests corrective or preventive measures. The method is proved to be very accurate in predicting the fault condition based on sensed parameter values. In addition, this method allows reduction in the size of the transmitted data packets to 23 bits, which help in increasing the overall network lifetime of the WSN system deployed in the wind farm.

Acknowledgements

I would like to thank Amity University, Lucknow, for providing me the means and resources in writing this chapter. I sincerely thank our Pro Vice Chancellor and Director ASET, Lucknow campus..

Author details

Deepshikha Agarwal

Address all correspondence to: sun.mita@gmail.com

Amity University, Lucknow, India

References

[1] Wiser RH, Bolinger, M. Wind Technologies Market Report. Energy Efficiency & Renewable Energy, U.S. Department of Energy, Springfield, DOE/GO-102014-4459. August 2014. pp. 1-95.

[2] www.eia.gov/energyexplained. [Accessed on February 1, 2017]

[3] www.davidsuzuki.org/issues/climate-change. [Accessed on February 1, 2017]

[4] www.ucsusa.org/clean-energy/coal-and-other-fossil-fuels. [Accessed on February 1, 2017]

[5] Lau BCP, Ma EWM. Review of offshore wind turbine failures and fault prognostic methods. In: Proceedings of IEEE Conference on Prognostics and System Health Management; Beijing; May 23–25, 2012. pp. 1-5

[6] Hansen MH. How Hard Can it be to Pitch a Wind Turbine Blade? Silkeborg, Denmark: RISO Lab, Denmark Technical University; 2007

[7] Teknik FS. Advantages of Hydraulic Pitch Control. Available from: www.fst.dk/dyn/files/normal_items/1029-file-2/paper.Pdf [Accessed: August 15, 2016]

[8] Mohanty AR, Kar C. Fault detection in a multistage gearbox by demodulation of motor current waveform. IEEE Transactions on Industrial Electronics. June 2006;**53**(4):1285-1297

[9] Zhang Z, Verma A, Kusiak A. Fault analysis of the wind turbine gearbox. IEEE Transactions on Energy Conversion. 2012;**27**:526-535

[10] Brandão RF, Carvalho JA, Maciel Barbosa FP. Forecast of faults in a wind turbine gearbox. In: ELEKTRO 2012. Rajeck Teplice, Slovakia.

[11] Badihi H. Model-based active fault-tolerant cooperative control in an offshore wind farm. Energy Procedia. December 2016;**103**:46-51

[12] An M-S. Implementation of fault diagnosis of wind turbine based on signal analysis with NN algorithm. In: International Conference on Disaster Recovery and Business Continuity (DRBC); Korea; November 25-28, 2015.

[13] Rajagopalan S, Restrepo JA, Aller J, Habetler T, Harley R. Non stationary motor fault detection using recent quadratic time frequency representations. IEEE Transaction on Industry Applications. June 2008;**44**:735-744

[14] Blodt M, Bonacci D, Regnier J, Chabert M, Faucher J. Online monitoring of mechanical faults in variable-speed induction motor drives using the Wigner distribution. IEEE Transaction on Industrial Electronics. February 2008;**55**:522-533

[15] Riera-Guasp M, Antonio-Daviu J, Roger-Folch J, Palomares MPM. The use of the wavelet approximation signal as a tool for the diagnosis of rotor bar failure. IEEE Transaction on Industry Applications. June 2008;**44**:716-726

[16] Amirat Y. Wind turbines condition monitoring and fault diagnosis using generator current amplitude demodulation. In: 2010 IEEE International Energy Conference, Bahrain, December 2010

[17] Al-Ahmar E, Benbouzid MEH, Amirat Y. DFIG-based wind turbine fault diagnosis using a specific discrete wavelet transform. In: International Conference on Electrical Machines, Portugal, September 2008

[18] Agarwal D, Kishor N. Network lifetime enhanced tri-level clustering and routing protocol for monitoring of off shore wind farms. IET Wireless Sensor Systems. 2014;**4**:69-79. DOI: 10.1049/iet-wss.2013.0050

[19] Agarwal D, Kishor N. Flexible threshold selection and fault prediction method for health monitoring of off-shore wind farm. IET Wireless Sensor Systems. 2015;**5**:183-192. DOI: 10.1049/iet-wss.2014.0008

[20] Agarwal D, Kishor N. A flexible threshold selection and fault detection method for health monitoring of offshore wind farms. In: 2nd International Conference on Information and Communication Technology; Bandung, Indonesia; May 28-30, 2014. 2014. pp. 493-498

[21] Agarwal D, Kishor N. A fuzzy inference based fault detection scheme using adaptive thresholds for health monitoring of offshore wind farms. Special Edition on IEEE Sensors Journal. 2014;**14**:3851-3861. DOI: 10.1109/JSEN.2014.2347700

Edited by Mustafa Demetgul and Muhammet Ünal

Mass production companies have become obliged to reduce their production costs and sell more products with lower profit margins in order to survive in competitive market conditions. The complexity and automation level of machinery are continuously growing. This development calls for some of the most critical issues that are reliability and dependability of automatic systems. In the future, machines will be monitored remotely, and computer-aided techniques will be employed to detect faults in the future, and also there will be unmanned factories where machines and systems communicate to each other, detect their own faults, and can remotely intercept their faults. The pioneer studies of such systems are fault diagnosis studies. Thus, we hope that this book will contribute to the literature in this regard.

ISBN 978-953-51-4822-7

Photo by v_alex / iStock

IntechOpen