University of Alberta

Fault Detection and Diagnosis in Nonlinear Systems, with a Focus on Mining Truck Suspension Strut

by

Mohammad Hajizadeh

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

Department of Mechanical Engineering

©Mohammad Hajizadeh

Spring 2014

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the university of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

ABSTRACT

Classical fault detection methods do not completely satisfy the reliability requirement for complex and highly nonlinear stochastic systems. One solution to this problem is to use more advances fault detection methods such as multiple models to simulate system in different operating conditions.

This study focuses on fault detection and identification (FDI) of suspension strut and particle filter is used as estimator in interacting-multiple-model-based (IMM-based) structure. The main idea of the IMM-based diagnosis algorithm is that the actual system is assumed to have uncertain (failure status) parameter vector affecting the matrices defining the structure of the model. Then, a model set is defined to model each of these different parameters and each model is in certain probability drawn from model set. By calculating these probabilities one can determine the mode in effect at each sampling time and perform fault detection and diagnosis and determine the presence of a particular failure mode.

Acknowledgements

I would like to thank my advisor, Professor Mike Lipsett, for all the time, support, and guidance he has provided throughout this process. An additional thanks goes to the professors and staff of the Department of Mechanical Engineering at University of Alberta who have taught me so much and made my experience here so wonderful.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation and Statement of the Problem

A fault is an undesirable state of a system which affects system function. It may reduce productivity in the system, or in severe case may lead to breakdowns and catastrophes with serious consequences (e.g. Chernobyl or Bhopal disasters). Systems at risk of failure may need high level of reliability or may not be economical to modify them, but they may be monitored, in which case faults must be accurately detected and identified. There is a need for improved fault detection and identification (FDI) systems to continuously evaluate the condition of the monitored system without interrupting operation, by recognizing anomalous behavior due to system fault.

Furthermore, in many systems, performance requirements continually increase, resulting in a high degree of sophistication and automation, with corresponding complexity. Also, because of ever-increasing demand for safety, reliability and maintainability as well as more strict environmental legislations, there is a growing need for development of more reliable and accurate systems.

FDI methods have been well developed over the past three decades, as a field of study (Ding, 2007; Gertler, 1998; Isermann, 2006) and applied to several industrial applications (Mehra, 1998; Rigatos, 2009; Wang and Syrmos, 2008).

Most research on FDI has focused on linear systems and fault detection and identification for nonlinear systems has not been investigated to a great extent. For nonlinear systems, a general framework does not exist, the FDI methods for nonlinear systems have been developed based on specific assumptions and they have their own limitations. Some of these methods can address specific kind of nonlinearities or can just apply to systems working around particular operating point. For stochastic systems, much of development in fault detection schemes assumes that noises and disturbances are Gaussian. These fault detection methods do not completely satisfy the reliability requirements for complex and nonlinear systems with wide operating range and cannot meet customer's needs and expectations. Thus, developing an integrated and more reliable FDI solution, that can perform fault detection and identification for wider range of nonlinear stochastic systems, is highly desirable and necessary. Nonlinear FDI methods are usually complex but they can be facilitated by faster computers and also extensive number of sensors collecting different types of signals related to information on equipment operating condition.

A good example of modern complex systems is a mining haul truck (Figure 1.1). This heavy industrial equipment works in very harsh environments and it demands a high degree of reliability to keep the productivity of the mining operation (Bongers and Gurgenci, 2008). Harsh and highly variable operating conditions pose many reliability challenges for trucks and they are prone to a variety of faults. Off-road haul trucks are usually driven on rough, unpaved roads and on mining benches for loading, hauling, and on soft uneven ground to dump overburden and gangue material. A significant amount of energy is translated into vertical motion in these trucks, due to these uneven ground and road bumps. This undesirable vertical motion is dampened by the vehicle suspension system that must be responsive and tough. Strut is the key element in vehicle suspension system and incurs continual motion during harsh operating conditions (Figure 1.1). This ongoing motion causes strut performance to degrade over time, due to wear of the seals and loss of hydraulic fluid, or by failure of the control valves in strut. A single collapsed strut can cause serious structural damage and tire wear. Tires are a major operating expense of off-high-way trucks. Although currently preventive maintenance is being used for struts, strut failure rates are not predictable. Therefore an automated fault detection system to detect and identify anomalies in the struts would improve fleet reliability significantly.

There are some techniques developed for anomaly detection and diagnosis for suspension system. FDI applications to strut problem have been reported in both automobile industry (Börner et al., 2000; Fisher et al., 2003; Isermann et al., 2002; Majjad, 1997) and railway industry (Goda and Goodall, 2004; Goodall, 2006; Li and Goodall, 2004; Mei and Ding, 2007). Most of the studies are concerned with model-based techniques which use mathematical models to generate additional output signals and compare it with the original measurable parameters; however, these researches have been based on approximate linear suspension dynamic models and they usually not consider fault diagnosis. Strut in mining haul truck is a nonlinear system with different types of faults and it operates in a harsh operating condition with multiple operating points and there is a need to develop an FDI method that can handle nonlinearity and different operating conditions and noises in mining truck's strut.

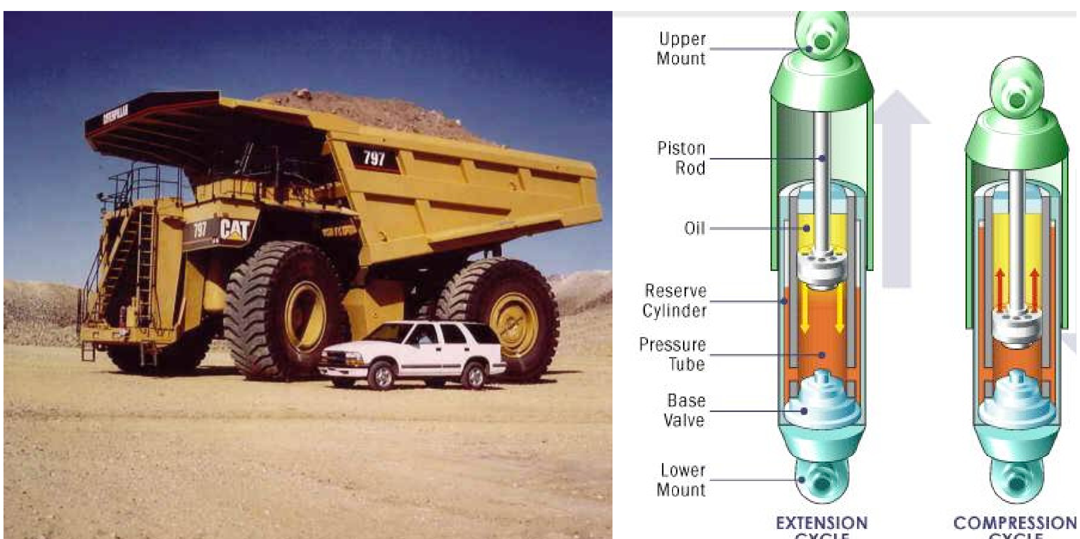

Figure 1.1: Mining haul truck and strut suspension (http://www.howstuffworks.com)

1.2 Objectives

The main objective of this study is to propose an FDI technique for truck suspension strut as a nonlinear dynamical system. This study will focus on the observer-based method, because of its abilities to deal with nonlinear systems. The idea of observer-based fault detection is to generate estimations of measured signals using a model of the monitored process, and compare the measurements with their estimations so as to generate a symptom signal that carries diagnostic information about system faults. However, the presence of noises and disturbances is inevitable. Therefore, the aim is to design observers such that the effect of noises and some disturbances on the residual signal is reduced, while the effect of a faults is considerably increased. This study provides a framework for an automated fault detection and diagnosis system for a mining truck suspension strut as a nonlinear system with the aim of reducing fault detection time, false alarm rates, undetected faults and unscheduled maintenance downtime. The FDI scheme should be able to reliably detect the presence of anomalies in strut with presence of non-Gaussian noises and isolate the problem, and identify the type of fault.

1.3 Organization

This study provides a brief literature review about available fault detection methods and then it gives a detailed explanation about interacting multiple model fault detection approach. Lately, the application of this specific approach is demonstrated on mining truck suspension strut as the application area for which this research is targeted. Following the first chapter of this thesis, a review of literature and relevant background in fault detection methods and more specifically model based fault detection methods is presented in Chapter 2. Fundamental theories of interacting multiple model (IMM) fault detection method are explained in Chapter 3. Application of particle filter in IMM structure for fault detection and diagnosis is discussed in Chapter 4 and algorithm for application of the particle filter is then provided. The results of application of data-driven methods for fault detection and diagnosis of suspension strut as a preliminary work showing limitations of conventional fault detection and isolation (FDI) methods for suspension strut are presented in Chapter 5 followed by the results of application of IMM-based FDI method on hydraulic two tank problem as a bench mark problem and suspension strut problem. Finally, Chapter 6 discusses future applications of this work as well as conclusions from this study.

Chapter 2: Literature Review

The task of fault detection and isolation is to determine whether there is a fault or not and to identify the fault, from the observation and the knowledge of the process. The international federation of automatic control (IFAC) technical committee on fault detection supervision and safety of technical processes (SAFEPROCESS) standardized the definition of the terminology used in fault diagnosis. These definitions are as follows:

Fault: A deviation of at least one characteristic property or parameter of the system from the acceptable, usual, or standard condition.

Failure: A permanent interruption of a system's ability to perform a required function under specified operating conditions in a way that it has to be shut down.

Fault Detection: Determination of faults present in a system and the time of fault occurrence.

Fault Isolation: Determination of the kind, location, and time of the occurrence of a fault.

Fault Identification: Determination of the size and the time-varying behavior of a fault.

Fault Diagnosis: Determination of the kind, size, location, and the time of the occurrence of a fault. It includes fault detection and identification.

Monitoring: A continuous real-time task of determining the conditions of a physical system, by recording information, recognizing and indicating anomalies in the system behavior.

Linear and nonlinear systems: Definition of the terminology used in this report are as follows:

Linear systems satisfy the properties of superposition and homogeneity. The principle of superposition states that for two different inputs, x and y, in the domain of the function f,

$$f(x + y) = f(x) + f(y)$$

The property of homogeneity states that for a given input, x, in the domain of the function f, and for any real number α,

$$f(\alpha x) = \alpha f(x)$$

Any function that does not satisfy superposition and homogeneity is nonlinear.

Stochastic system: Systems contain some element of random or stochastic behavior are called stochastic systems. Unlike a deterministic system, a stochastic system does not always produce the same output for a given input, hence they are unpredictable and they don't have a stable pattern or order.

Operating point: the operating point is a specific point within the operation characteristic of a dynamic system and it defines system's overall state at a given time. Properties of the system and the outside influences and parameters determine the operating point of a system.

Sensor fault: sensor fault can be observed as measurements that are unavailable, incorrect or unusually noisy.

Actuator fault: an actuator fault corresponds to the variation of the control input signal applied to the system or a problem in the actuator such as stuck control valve. The actuator faults can be defined as any abnormal operation of any element in the actuator subsystem such that the control command from the controller output cannot be delivered to manipulated variables entirely.

Component fault: Structural changes or changes in the process itself which occur due to hard failures in equipment are called component fault."

First-order Markov process: A random process that has *n* possible states when it is moving through time is considered to be a first-order Markov process if the state at the next time period is only reliant on the current state of the system.

Gaussian and non-Gaussian noise: When frequency distribution of the noise in the system obeys a Gaussian distribution it is called Gaussian noise, otherwise it is called non-Gaussian noise.

Once the fault detection and diagnosis has been done, it is possible to make a decision on the action to be taken. The whole process -fault detection, diagnosis, evaluation and action- forms a supervision system (Isermann, 1984), as shown in Figure 2.1.

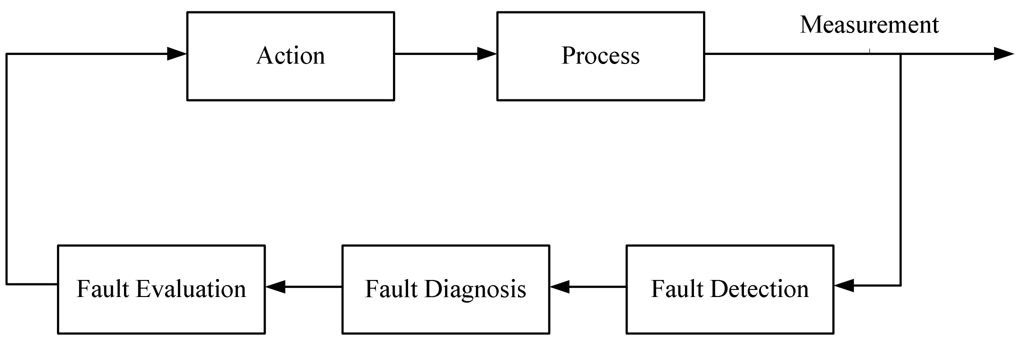

Figure 2.1: Supervision loop (Li, 1991)

The earliest fault detection methods simply monitored a particular signal and raised an alarm when it exceeded a predefined value. A more advanced technique used to identify faults in the system was hardware redundancy, which entails replicating the hardware (such as actuators and sensors) in the plant and running it in parallel with the current hardware. Any discrepancy between the channels will indicate a fault. The major drawback of hardware redundancy is the cost of replicating hardware, which is often prohibitively expensive and in some applications this method may not be physically possible. Research on fault detection and isolation first began in the early 1970s by Beard (1971) and Jones (1973), who considered analytical redundancy instead of hardware redundancy. The analytical redundancy-based approach is a health monitoring software that calculates the differences between estimates and measurements of the system states and generate residuals using a system process model. Analytical redundancy-based approaches can be more applicable and cheaper compared to hardware redundancy techniques, because the hardware are not replicating and also no additional faults are introduced into the system. They require an accurate model, and significant computing power and data storage (Bhagwat et al., 2003; Yen and Ho, 2003).

Analytical fault detection methods can be commonly categorized in three main classes, namely model-based, data-driven, and Knowledge-based fault detection (Ding, 2008; Venkatasubramanian et al., 2003a; Venkatasubramanian et al., 2003b; Venkatasubramanian et al., 2003c; Chen and Patton, 1999; Patton et al., 1998). The method employed is, however, dependent on the type and nature of system unit to be monitored and information that we can get from the system through sensors and also knowledge that we have about the system. All these approaches generate characteristic signals or information that can be used to monitor the system. Brief explanations of these methods are given below:

- Data-Driven – In data-driven approaches, it is unnecessary to assume the availability of models and only the availability of a large amount of historical process data is assumed. This method for fault detection is based on analysis of the measured output signals. The measured output is analyzed or filtered to yield features and further information regarding the detection of faults. These features can be in time or frequency domain, some examples are the signal mean, variance, skewness, kurtosis, crest factor or the power in a certain frequency band.
- Knowledge-Based – Sometimes a process is too complex to be modeled analytically and also a regular data-driven fault detection approach is not reliable. In that case, knowledge from past experience or process history can be used to develop an evidential process model for fault detection. This technique is based on a qualitative model of the system. Qualitative process

knowledge can be used to evaluate relations between measured signals and the current operating conditions and to classify faulty behavior of the system. Different techniques of knowledge based methods are available such as: causal analysis, expert systems, pattern recognition, etc.

- Model-Based – Fault conditions are determined from deviations between a theoretical model and the physical process. Different approaches for model-based fault detection have been developed. These approaches use first-principles knowledge to develop a mathematical model of the system and to calculate process quantities such as outputs, parameters or states, which are called features. By comparing these features from the model with actual values from the process, and analyzing the differences or residuals, symptoms can be generated that can be used in fault detection and diagnosis.

Data-driven techniques for FDI do not require model information because they rely on data measured from the system, and take advantage of signal processing and pattern classification techniques. These techniques are often used to detect discrete changes in a process (Freeman et al., 2013). To identify patterns that classify normal or fault conditions, the time-domain and frequency-domain signal processing techniques may be used (Zhou et al., 2010). The main advantage of data-driven techniques is their ability to transform the high dimensional data into a lower dimension, without losing any important information (Chiang et al., 2001). Data-driven approaches are generally classified as non-statistical (e.g. neural networks (NN)) or statistical (e.g. principal component analysis (PCA), partial least squares (PLS) and statistical pattern classifiers) methods (Yang, 2004).

The use of PCA and PLS in process analysis and fault diagnosis was reviewed by MacGregor and Kourti et al. (1995) and Wise and Gallagher (1996). Nomikos and MacGregor (1994) applied multivariate projection methods to batch processes using multi-way PCA. To handle nonlinearity, Qin and McAvoy (1992) proposed a combined approach of a neural network PLS method, in which the PLS modeling integrated with feedforward networks. Dong and McAvoy (1996) used a nonlinear PCA method to deal with nonlinearity. Raich and Cinar (1996) proposed an integrated statistical methodology that integrated PCA with discriminate analysis techniques, using distance-based and angle-based discriminants. Moreover, Dunia et al. (1996) and Qin and Li (1999) utilized PCA for sensor fault detection, identification and reconstruction. Most real processes are time-varying; but a PCA model is time invariant; and so it should be updated. To overcome this problem Qin (1998) and Li et al. (2000) proposed an adaptive monitoring approach using recursive PLS. Luo et al. (1999) and Zhang et al. (1999) proposed multiscale PCA (MSPCA) as another variant of the PCA approach which integrates PCA and wavelet analysis. Later in 2008, Hongxing et al. (2008) presented the enhancement of fault detection by combining PCA and wavelet packet transformation. Wu and Huang (2008) developed a kernel PCA method integrated with a wavelet transform to modify fault monitoring in a nonlinear chemical process. To enhance the fault detection accuracy, Xia et al. (2008) proposed a new algorithm based on the kernel PCA and wavelet packet transform. In addition, Ning et al. (2010) developed a novel PCA method integrated with wavelet denoising and the results showed that wavelet denoising improved the accuracy of fault detection.

Data-driven methods can be used to extract fault information from signals measured in nonlinear systems but they have their limitations. The main downside of data-driven techniques is that numerous false alarms are generated by the fault detection system, because these techniques do not consider the dynamic interrelationship between the measured signals of the system. Wavelet transform integrated with PCA has been applied on haul truck suspension system for fault detection (as described in Chapter 5). Results showed that this data-driven method can detect the fault in this system; but, because of nonlinear nature of the process, this method shows poor result in some parts of the studied data. The other problem is that there are some disturbances and noise in the system that cannot be modeled as a Gaussian signal; and so there is a need for more sophisticated FDI method to handle non-Gaussian noises and disturbances in the system.

Knowledge-based methods use neural networks or fuzzy system techniques to handle nonlinearities in the systems. Some typical problems in these methods include the network selection, the training algorithms and the representation of system dynamics by neural networks or fuzzy system. Another problem is that they do not give insight into the physical model of the system, therefore the output of their model does not have physical meaning and it is hard to interpret them and diagnose the fault.

If the dynamic behaviour of a system can be well described by an accurate mathematical model, then analytical model-based methods are the most powerful fault diagnosis methods (Li and Kadirkamanathan, 2001).

The general procedure of model-based FDI is as follows. 1) A residual is generated from the difference between real measurements or system parameters and estimates of these measurements or parameters by a mathematical model. 2) After residual generation, the residual is evaluated by extracting features from the residual. A common approach is simple threshold checking. 3) Finally, the decision making step analyzes the result of the evaluation of a set of residuals by using numerical and statistical techniques for the likelihood of faults. It returns a decision as to which fault(s) have occurred and which component is faulty in the supervised system.

Once fault detection and diagnosis has been done, it is possible to make a decision on the action that needs to be taken depending on the type and extent of the fault. Figure 2.2 illustrates the general and conceptual structure of a model-based fault diagnosis system.

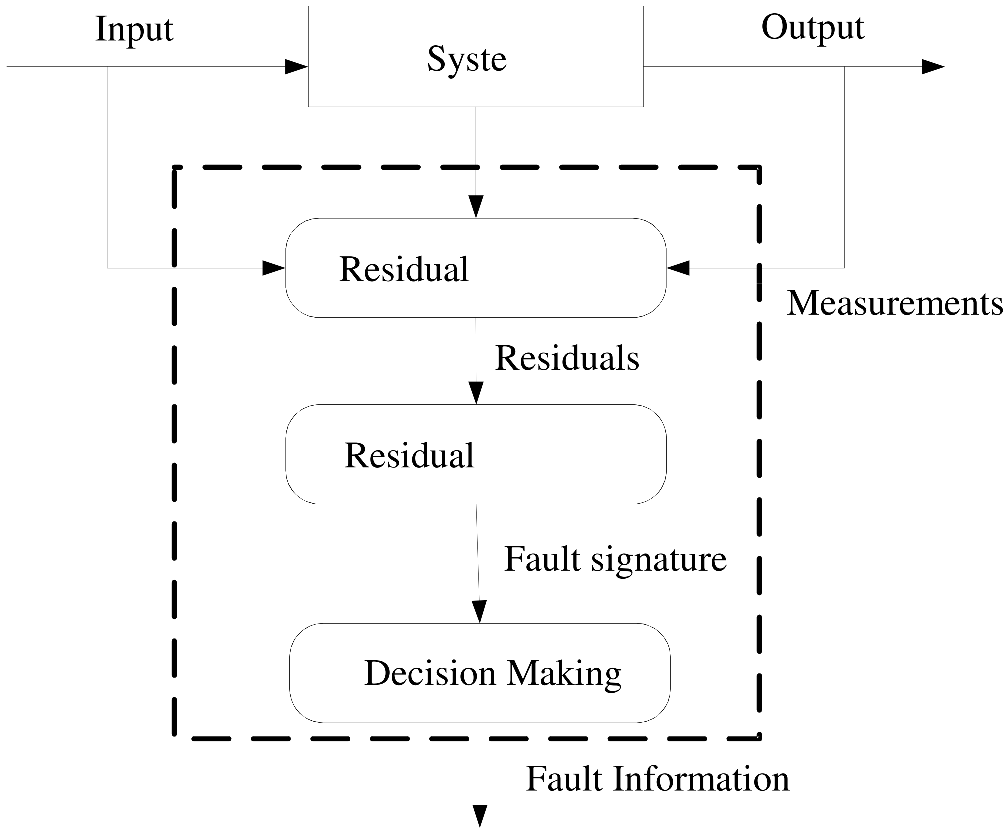

Figure 2.2: General Structure of Model-Based Fault Diagnosis Systems (Yang, 2004)

In general, three main model-based approaches are used to generate residuals (Freeman et al., 2013). The first is parameter estimation (Isermann, 1994). This approach is based on the assumption that the faults are reflected in some parameters of the system, often related to physical parameters. Hence, in order to identify faults, the system parameters are estimated on-line using well-known parameter estimation techniques. The residuals in this approach are essentially the difference between the on-line estimates of the system parameters and their corresponding values under fault-free conditions.

The second is the parity space approach (Chow and Willsky, 1984), where the residual is generated using so-called parity functions defined over a time window of system input and output data. The parity space method is based on simple algebraic projections and geometry and the basic idea behind this method is to provide an appropriate check of the parity or consistency of the various measurements within the monitored system.

Finally, the third approach to residual generation is the observer-based methods (or filter-based, or state estimation-based) (Chen and Patton, 1999).

Kinnaert et al. (2000) designed a model-based diagnosis system to detect actuator, sensor and component faults in a gas-liquid separation unit and validated it by simulation. Thumati and Jagannathan (2010) developed a novel model-based fault detection and prediction (FDP) method for nonlinear multiple-input–multiple-output (MIMO) discrete-time systems. The performance of the proposed method is demonstrated by a fourth-order MIMO satellite system (Thumati and Jagannathan, 2010).

Model-based FDI based on neural network and fuzzy methods were applied by several researchers. For example, a new structure of partially connected neural networks (PCNN), and a conventional, fully connected neural network (FCNN) was used as identification techniques for nonlinear systems to generate residuals for fault detection purposes (Fekih et al., 2007).

Gertler et al. (1993) used the parity equation residual generation method for fault diagnosis of systems modeled by a linear dynamic core with static input and output nonlinearities (hybrid or Hammerstein model). Krishnaswami et al. (1995) used the nonlinear autoregressive moving average modeling with exogenous inputs (NARMAX) method for system identification to apply a nonlinear parity equation residual generation (NPERG) FDI scheme, for diagnosing faults in an automobile engine.

Although there are some application of parity space method to nonlinear systems, it should be noted that this method sensitive to measurement noise and process noise (or disturbance) because it does not have closed-loop structure like observers or filters (Sobhani-Tehrani, 2008).

There are some researches on the application of parameter estimation method for nonlinear system. The most important issue for parameter estimation based diagnosis for nonlinear systems is model complexity. It is difficult to use this method in real time application because parameter estimation for the nonlinear model is a multivariate nonlinear optimization problem (venkatasubramanian et al, 2003b).

The basic observer-based approach is to reconstruct the outputs of the system from the measurement or subset of measurements by using observers or filters, and defining the residuals as the difference between the actual measurements and the model-based estimates. This residual can be used for fault detection. Observer-based methods are one of the most popular approaches to model-based FDI (Isermann and Balle, 1997) and various observer-based fault-detection methods have been developed (Frank, 1987) and each of them has certain merit and applicability.

Observers are often used in control systems to estimate non-measurable states (which is usually the case in health monitoring systems). An observer is essentially a system model, and can fall into one of two categories (Samy and Gu, 2011, Alrowaie et al., 2012): a Luenberger observer (used in a deterministic setting) or a Kalman filter (used in a stochastic setting). The estimation errors of the Luenberger observer (often simply referred to as observer) or the innovation sequence of the Kalman filter can be used as residuals for FDI purposes. Figure 2.3 shows the schematic diagram of an observer-based residual scheme.

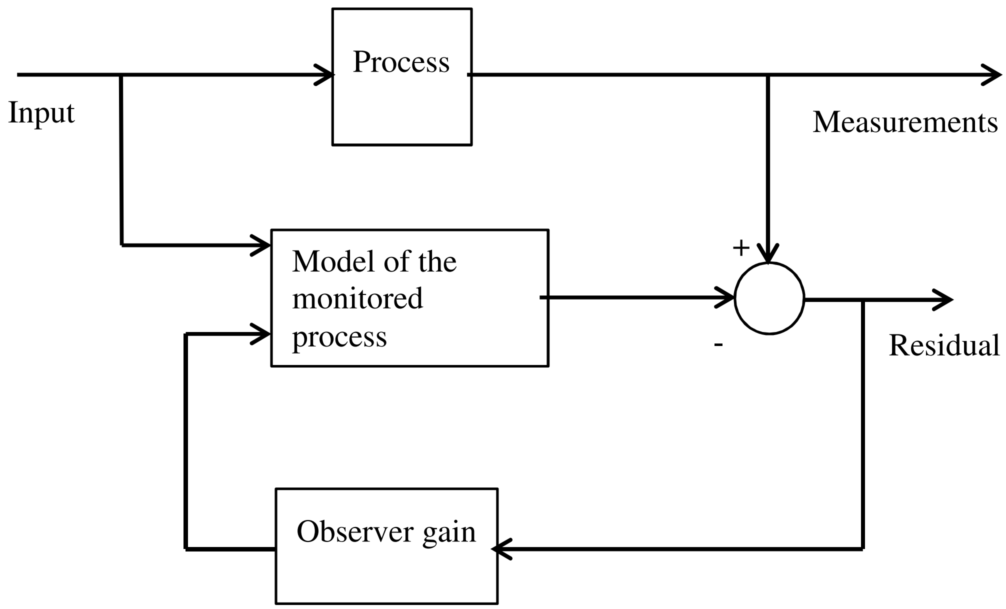

Figure 2.3: Schematic diagram of observer-based residual generator

Residual generation schemes that utilize the theory of nonlinear observers were proposed in several studies (Hengy and Frank, 1986; Himmelspach, 1992; Seliger and Frank, 1991). A nonlinear observer estimates the output of the system to calculate the residual for fault detection (Bhagwat et al., 2003b). Zolghadri et al. (1996) applied two nonlinear observer-based FDI schemes to a laboratory hydraulic installation under digital control. The practical results demonstrated that faults can be successfully detected and isolated using both methods. Pertew et al. (2007) proposed new LMI (linear matrix inequality) observer design for Lipschitz nonlinear systems which offers extra degrees of freedom compared to classical Lipschitz observers. The proposed observer applied for sensor fault detection and diagnosis in nonlinear Lipschitz systems.

The kalman filter (KF) is the most broadly used and popular filter for linear stochastic systems. Mehra and Peschon (1971) used the residuals of KF for fault detection. Isermann (1984) developed an algorithm using KF state estimation techniques for pipeline leak detection. The EKF is the linearized form of the KF that can be used for nonlinear systems (Bhagwat et al., 2003a). Dalle Molle and Himmelblau (1987) proposed FDI method by parameter estimation using EKF for evaporator. Li and Olson (1991) used EKF for FDI of a closed-loop simulated nonlinear distillation process. For real-time applications, the use of a single filter for all measurements and models is inconvenient and usually inapplicable (Bhagwat et al., 2003a). In order to solve this problem several techniques have been suggested. For example, Gelb (1974) simplified the model equations by decoupling them and using separate local EKF. Fathi et al. (1993) designed separate local EKFs based on models of each component instead of a single complex filter. Also Bhagwat et al. (2003a) used the separate filters to overcome this problem during practical implementation.

Since representing all possible system modes with a single model is often impractical, the use of multiple model-based approaches for modeling and control purposes has become popular. The multiple model (MM) approach is a very powerful tool for estimation, control and modeling problems and it has attracted considerable research interest in the recent decades (Li and Bar-Shalom, 1993, Willsky and Jones 1974). It has been used in target tracking problem and also fault detection and diagnosis scheme for systems with different kind of faults (Willsky, 1976). Banerjee et al. (1997) reported the challenges in state estimation of a nonlinear plant that operates in multiple modes and makes transitions between them. To overcome this problem they identified local linear models at each operating points and used them to develop a control system. Johansen and Foss (1999) introduced multiple model-based techniques that decompose a broad range of processes operating regimes into a set of separate operating regimes and develop local models. Also, Sun and Hoo (1999) employed multi-linear method and proposed a dynamic transition controller to control the single-input single-output (SISO) nonlinear processes. The performance of the method was demonstrated by simulation of pH neutralization process. Galan et al. (2000) also used a multi-linear model to design H-infinity control of nonlinear plants. This method was applied on a bench-scale pH neutralization continues stirred-tank reactor (CSTR).

To apply multiple model-based (MM) approaches for FDI, a model set containing local models corresponding to different fault conditions and normal conditions of the system may be developed. A multiple model contains a bank of filters based on the model of each system mode which works in parallel. The overall state is estimated by weighted sum of the outputs of all filters based on the probability of occurrence. Bhagwat et al. (2003b) proposed a model-based fault detection method that decomposes the nonlinear transient systems into multiple linear modeling regimes. Kalman filters and open-loop observers were used for state estimation and residual generation based on the resulting linear models. The method was validated against experimental data obtained during the startup transition of highly nonlinear pH neutralization reactor in the laboratory.

In MM-based approaches, the filters are running in parallel without mutual interaction (that is, “non-interacting” MM estimation). This method is effective for a systems without structural or parametric (mode) changes (Raghappriya et al., 2012). To overcome this downside of a non-interacting MM approach, interacting multiple model (IMM) approach, that handles abrupt changes in the dynamic of system by switching from one model to another model in a probabilistic manner, was developed. The IMM estimator was first introduced by Bar-Shalom and Fortmann (1988), and then Efe and Atherton (1997) applied it to fault detection (Donders, 2004). Thereafter, some studies have been conducted on IMM-based FDI which implemented along with different filters such as KF, EKF and UKF for estimating states and mode probabilities.

Hayashi et al. (2006) applied an IMM algorithm for fault detection of railway vehicles. The KF was used to estimate the mode probabilities and states of vehicle suspension. The simulation results indicated that the algorithm is effective for railway vehicle suspension fault detection. Gadsden and Habibi (2011) used IMM algorithm for fault detection and diagnosis of hybrid electric vehicle (HEV) battery system. The smooth variable structure filter (SVSF) was used instead of the KF in this algorithm. Gadsden (2012) also reported successful application of IMM-SVSF method for fault detection and diagnosis of an electrohydrostatic actuator (EHA). Raghappriya et al. (2012) proposed an IMM algorithm for detection and diagnosis of multiple faults. The reliability of the proposed method was illustrated by an aircraft example with multiple failures of sensors, actuators, and other component failures.

For the nonlinear system FDI problem, Extended Kalman filter (EKF) is usually applied instead of Kalman filter (Wang and Syrmos, 2008). Mehra et al. (1998) proposed a novel IMM-EKF approach for FDI in nonlinear systems. The basic idea is to describe each failure mode by a separate model and to combine the outputs of EKFs based on different models in a near-optimal way. This approach was applied to a problem of spacecraft autonomy for FDI of sensor and actuator failures. The results demonstrated that this IMM-FDI filter successfully manages both permanent and transient failures. Kadirkamanathan et al. (2001) also used an IMM estimator that includes a number of EKFs running in parallel. The results of simulated and real data showed that IMM estimator is an effective method for tracking rapid trajectory changes. Wang and Syrmos (2008) applied the IMM-EKF algorithm for the detection and diagnosis of sensor and actuator faults with the purpose of condition monitoring of the electro-hydraulic actuator (EHA) system. The results demonstrated that IMM estimation algorithm yields more robust detection and estimation.

IMM-based method based on the Unscented Kalman filter (UKF) can be used for FDI in case of complex systems. Tudoroiu et al. (2006 and 2007) applied IMM method based on the UKF algorithm for FDI of partial (soft) and total (hard) faults of the satellite attitude control system's actuators. The method was implemented based on a high-fidelity and highly nonlinear model of a commercial satellite attitude control system. The results of numerical simulations showed that the method is more accurate, less computationally demanding, and more robust with the potential of extending to a number of other engineering applications compared to other model-based fault detection, diagnosis and isolation strategies.

When the system is highly nonlinear and the noises are non-Gaussian, the performance of EKF and UKF decreases, and may even it becomes divergent (Ristic et. al., 2004; Souibgui et al., 2011). Handschin and Mayne (1969) proposed the use of Monte Carlo simulation techniques for non-linear non-Gaussian state estimation but it was not popular due to high computational load. Recently due to increase in computational power of modern computers, the Monte Carlo filter has become popular (Li and Kadirkamanathan, 2001; Li and Kadirkamanathan, 2004). Sequential Monte Carlo (SMC) methods, also known as particle filtering (PF), have gained particular attention (Kadirkamanathan et al., 2001; Orchard and Vachtsevanos, 2009). The main advantage of particle filters is that they can handle any functional nonlinearity and system or measurement noise of any probability distribution (Kadirkamanathan et al., 2001). Unlike a conventional EKF-based approach that uses only the mean and covariance of an approximate Gaussian distribution, PF approach utilizes the complete probability distribution information of the state estimates for fault detection (Li and Kadirkamanathan, 2001; Kadirkamanathan et al., 2002). PFs approximate the required probability density function (PDF) by groups of particles in the state space. As the number of particles increases, they effectively provide a good approximation to the required PDF (Chen et al., 2005). The major drawback of PF is the computational load associated with the number of particles used. Thus it is often necessary to compromise between computation time and the quality of results, especially in the system with limited computational resources such as mobile robots (Zajac, 2011). Rigatos (2009) evaluated the performance of the PF against EKF for mobile robot fault diagnosis. The results demonstrated the effectiveness of PF.

Li and Kadirkamanathan (2001) developed a PF based MM approach for fault detection and identification by combining PF with Bayesian inference. They showed the effectiveness of the proposed method by experimental results. Kadirkamanathan et al. (2002) also developed another PF based MM approach for FDI by combining PF with the likelihood ratio (LR) test. The simulation results on a highly nonlinear system exhibited the effectiveness of the proposed method. Li and Kadirkamanathan, in their previous works, assumed that the knowledge on all possible faults of the system is available, and each of the possible faults can be described by a known model. Later in 2004, they modeled faults as the unknown changes in the parameters of the system and proposed the adaptive Monte Carlo filters which integrated with the likelihood ratio test (Li and Kadirkamanathan, 2004). Also Souibgui et al. (2011) combined PF with the innovation based fault detection techniques to develop a fault detection and isolation scheme. The simulation results on a highly nonlinear system exhibited the success of the PF. Alrowaie et al. (2012) proposed the algorithm to modify the log-likelihood ratio test presented in Kadirkamanathan et al. (2002) and Kadirkamanathan and Li (2004) as it derives the likelihood function through an approximation that is not applicable to all types of nonlinearities. Recently, Bruno (2013) also discussed Rao-Blackwellized particle filtering as a method that is suitable for many applications such as fault detection.

A summary of the pros and cons of some of discussed methods are summarized in Table 2.1.

Table 2.1: Summary of the pros and cons of the discussed methods

FDI Methods	**Pros**	**Cons**
Hardware Redundancy	• Accurate and fault tolerant.	• The cost of replicating hardware is too expensive. • May not be physically possible in some applications.

Data Driven	• Ability to transform the high dimensional data into a lower dimension, without losing any important information • Easy to apply	• Do not consider the dynamic interrelationship between the measured signals of the system, result in generation of numerous false alarms by the fault detection system. • Do not perform well in the presence of disturbances
Model Based	• The dynamic behaviour of system can be well-described by perfect mathematical models, thus the analytical model-based methods are the most powerful fault diagnosis methods. • Robust under the effect of noise and adaptable	• Requires greater computing power and data storage. • Difficult to apply using complex models
MM	• Effective when it is impractical to represent all possible system modes with a single model.	• Do not work well for the problems having large structural or parametric changes including failure or damage.
IMM	• Handle abrupt changes in the dynamic of system by switching from one model to another model in probabilistic manner • More accurate and promising for FDI scheme compared to MM algorithm because it considers interactions between different modes.	• Computationally expensive

IMM-KF	• Powerful and common tool for FDI.	• Do not work for nonlinear systems
IMM-EKF	• Powerful method for systems with simple nonlinearity working around single operating point.	• Poor performance when the system is highly nonlinear and the distribution is non-Gaussian
IMM-UKF	• More accurate, and more robust compared to other model-based FDI strategies. • Derivative-free alternative to EKF, and provides superior performance at an equivalent computational complexity.	• Poor performance when the system is highly nonlinear and the distribution is non-Gaussian
PF	• Utilizes the complete probability distribution information of the state estimates for FD. • Applicable to general non-linear systems with non-Gaussian noise and disturbances. • As the number of particles increases, they effectively provide a good approximation to the required PDF. • The computational complexity does not increase with increasing the number of states.	• High computational load associated with the number of particles used.

Chapter 3: Background on Multiple Model Fault Detection Methods

In this section basic concept of IMM-based FDI method is described and theoretical background and mathematical equations of this method are mentioned.

3.1 Multiple Model (MM)-Based FDI

Faults can be classified in three different classes according to their location, each yielding a different kind of models: sensor faults, actuator faults, and component faults (Witczak, 2007). Systems may be subject to changes due to different faults and variations caused by these changes may significantly alter the dynamic behavior of the system. Therefore, in model based fault detection, a suitable model of faults that represents the effects of a fault on the system is important and it is hard and sometimes infeasible to model all possible faults in the system with a single model. Each of these faults can be modeled with different model and all these model can work together to simulate the system in different modes. Consider a system that is subjected to a series of possible qualitative changes that make it switch, over time, among a countable set of models. Each of these models can be related to an operation mode or different fault of the system and the system jumps from one mode to the other. Then a filter is designed for each choice of system model, which results in a bank of single model-based filters. This bank of filters (or more precisely multiple models) can simulate continuous changes in the base states, and as well as jumps in system modes. They run in parallel, each based on a model matching a particular mode of the system or a particular fault. Usually one model in the bank is associated with the healthy operational mode and the rest of the models correspond to various possible fault scenarios in the system. However, multiple models associated with healthy operational mode can also co-exist in the bank, if the system structure changes during healthy operations.

The main idea of the MM-based diagnosis algorithm is that the actual system is assumed to have uncertain (failure status) parameter vector affecting the matrices defining the structure of the model. Then, a model set is defined to model each of these different parameters and parameters are assumed to take only discrete values to map the corresponding system models. Each system model is in certain probability drawn from models set designed to represent the all possible system behavior pattern. Then by calculating these probabilities one can determine the mode in effect at each sampling time. The whole model set is a hybrid estimator which estimates a state of a hybrid system. Hybrid states involve two types of components: those varying continuously (state of the conventional system), known as base states, and those that may jump only, known as modes or modal state. The diagnostic scheme uses the residual computed from comparing each of these estimator outputs to real model output to detect which fault is present. The residual is fed to the fault detection and identification module. This module uses the residual to calculate the probability of the different modes of operation and give out a warning signal to indicate the presence of a particular failure mode.

In MM algorithm each model operates independently without mutual interaction. Such an approach is quite effective in handling problems with various operating conditions but fixed structure or parameters. Whereas in reality the system structure or parameters do indeed change as a system component, a sensor or an actuator fails (Kim et al., 2008). Therefore, MM algorithm does not work well for the problems having large structural or parametric changes including failure or damage.

3.1.1 Interacting Multiple Model (IMM)-Based FDI

To make MM algorithms more suitable for FDI problem, new interacting multiple model-based FDI (IMM-FDI) approaches have been proposed (Mehra and Peschon, 1971). The IMM differs from the non-interacting MM algorithm in that the single model based filters interact with each other and thus result in improved performance. The IMM based estimator consists of a bank of single-model-based estimators running in parallel at each cycle as shown in Figure 3.1. The initial state estimates at the beginning of each cycle for each estimator are the mixture of all most recent estimates from single-based estimator. The IMM algorithm has more accurate state estimation compared to an MM algorithm and consequently more precise estimation of the posterior probability of each mode. Theoretical analysis and research results demonstrate that the IMM-FDI approach significantly improves the FDI performance in terms of fast detection and proper identification (Ru et al., 2008).

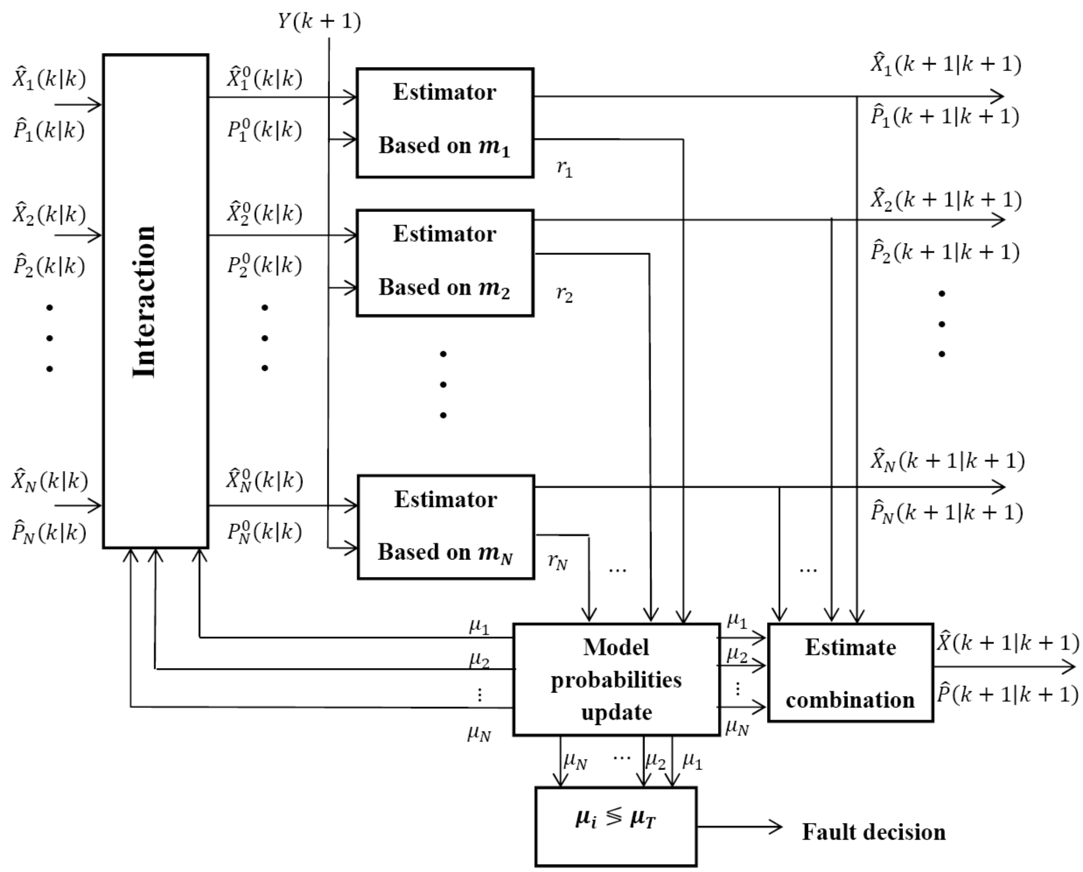

Figure 3.1: Block diagram of IMM-based FDI approach (Zhang and Li, 1998a)

In IMM modeling, the base state estimation is analogous to classical state estimation. The modal state estimation is, however, quite different. The IMM-based FDI approach assumes that the actual system at any time can be modeled sufficiently accurately by a stochastic hybrid system:

$$x(k+1) = f\big(x(k), u(k), m(k+1)\big) + w(k) \tag{1}$$

$$y(k) = h(x(k), m(k)) + v(k) \tag{2}$$

where $x \in R^{n_x}$ is the base state vector; $y \in R^{n_y}$ is the (mode-dependent) measurement vector; $u \in R^{n_u}$ is control input vector; $w \in R^{n_w}$ and $v \in R^{n_v}$ are mutually independent discrete-time process and measurement noises; $m(k)$ is the discrete-value that represents the current active system mode or model state at time k, and the set of all possible system mode is $\Theta = \{\theta_1, \theta_2, \ldots, \theta_s\}$.

It can be observed from the hybrid system defined in Equation (1) and (2) that system outputs are in general noisy and mode-dependent. Therefore, the mode information imbedded (i.e., not directly measured) in measurement sequences. The transition between the different models can be described as a first-order Markov process and transition probability between pairs of modes is denoted as:

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1s} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{s1} & \pi_{s2} & \cdots & \pi_{ss} \end{bmatrix} \tag{3}$$

and

$$\sum_j \pi_{ij} = 1, \qquad i = 1, \ldots, s \tag{4}$$

where π_{ij} is the transition probability from mode m_i to mode m_j.

In the IMM method, it is assumed that a set of N models has been set up to approximate the hybrid system (1)-(2) by the following N pairs of equations:

$$x(k+1) = f_j\big(x(k), u(k)\big) + w(k) \tag{5}$$

$$y(k) = h_j(x(k)) + v(k) \qquad j = 1, \ldots, N \tag{6}$$

where $N \leq s$ and subscript j denotes quantities pertaining to model $\theta_j \in \Theta$. System state and output equations may be different structures for different j.

Given that the probabilistic state-space formulation as Equation (5) and (6), the Bayesian approach provides a rigorous framework for the hybrid state estimation problem. In the Bayesian approach, the posterior probability density function (PDF) of the hybrid state $p(x_k, m_k|Y_k)$ is constructed based on all available information. If the posterior PDF is known, various estimates of the hybrid state, (i.e. base state and modal state) can be derived. The marginal distribution of the posterior PDF is derived as:

$$p(x_k|Y_k) = \sum_{i=1}^{N} p(x_k|m_k = m_i, Y_k)p(m_k = m_i|Y_k) \tag{7}$$

On the other hand, the fault diagnosis problem is to provide a distribution over the discrete set, Θ at each time step. This distribution can be obtained from the marginal distribution of the posterior PDF.

$$p(m_k|Y_k) = \int p(x_k, m_k|Y_k)\, dx_k \tag{8}$$

The above equation can determine the mode in effect at each cycle time. Observation is being updated at every time step, so new information can plug into the estimation process to make it more accurate. A recursive state estimation scheme can be designed to use new information in every time step. The posterior PDF, $p(x_k, m_k|Y_k)$, can be recursively estimated in two stages: prediction and update. The prediction stage involves using the state transition model to obtain the predicted PDF, $p(x_k, m_k|Y_{k-1})$, of the hybrid state at time k. The predicted PDF can be expressed as:

$$p(x_k, m_k|Y_{k-1}) = \int p(x_k, x_{k-1}, m_k|Y_{k-1})\, dx_{k-1} \tag{9}$$

where

$p(x_k, x_{k-1}, m_k|Y_{k-1}) = \sum_{m_{k-1}} p(x_k|x_{k-1}, m_k)p(m_k|m_{k-1})p(x_{k-1}, m_{k-1}|Y_{k-1})$

The probabilistic model of the state transition $p(x_k|x_{k-1}, m_k)$ is defined by the state-space Equation (5) and the known statistics of $w(k)$. $p(m_k|m_{k-1})$ and $p(x_{k-1}, m_{k-1}|Y_{k-1})$ are available from last iteration.

At time step k, a new observation *y(k)* becomes available and can be used to update the predicted PDF via Bayes' rule.

$$p(x_k, m_k|Y_k) = \frac{p(y_k|x_k, m_k)p(x_k, m_k|Y_{k-1})}{\int \sum_{m_k} p(y_k|x_k, m_k)p(x_k, m_k|Y_{k-1})\, dx_k} \tag{10}$$

Where $p(y_k|x_k, m_k)$ is the mode-conditional likelihood function defined by the measurement model (6) and the known statistics of $v(k)$ and $p(x_k, m_k|Y_{k-1})$ can be calculated from Equation (9). Equations (9) and (10) form the basis for the recursive hybrid state estimation scheme. Then $p(x_k, m_k|Y_k)$ can be used to find $\mu_i(k+1) = p(m_{k+1}|Y_{k+1})$, the estimation of the probability of mode *i*, at each time step.

By using the information provided by the model probabilities, both fault detection and diagnosis can be achieved. The fault decision can be made by

$$\mu_j(k+1) = \max_i \mu_i(k+1) \qquad i = 1, \dots, N \tag{11}$$

In this equation, we are looking for mode with the highest probability and μ_j is the probability of mode *j* which has the biggest probability among all mode probabilities, μ_i *i=1, … ,N*. A fault decision threshold, μ_T, can be defined and μ_j can be compared with this threshold. If μ_j is greater than the threshold, fault *j* is happened.

If μ_j

$\begin{cases} \text{greather than threshould then system is in mode } j \text{ and fault } j \text{ occured} \\ \text{otherwise no fault occured} \end{cases}$

It provides not only fault detection but also the information of the type, location, size and fault occurrence time, that is, simultaneous detection and diagnosis.

In IMM method, models are interacting with each other so at each time step k estimated posterior PDF for each model is the mixture of all posterior PDF from all models. Therefore, another stage called initialization stage should be added to prediction and update stage to perform interaction between the models. In this stage each filter is reinitialized by mixing the all most recent estimates from the single-model-based filters.

The probability of each mode in the next time step should be predicted as the initial value for next step state estimation. Based on the assumption that the transition between system modes is described as a first-order Markov process, the predicted mode probability for each mode $\mu_j(k+1|k)$ can be calculated from estimated probability of each mode by Bayesian algorithm and transition probability between the modes as:

$$\mu_j(k+1|k) = \sum_i \pi_{ij}\mu_i(k) \qquad i = 1, \dots, N \tag{12}$$

This recursive estimate scheme is only a conceptual solution and cannot be determined analytically. A method is needed to approximate the posterior state distribution ($p(x_k, m_k|Y_k)$) in the IMM algorithm. A recursive filter can be used to approximate the posterior state distribution. A model-based recursive filter has to be designed based on each model in the IMM model set in order to estimate system states. This model set generally consists of one model to represent healthy behavior of the system and multiple models with embedded fault model to represent different possible faults in the system. An increase in robustness is expected, because each filter uses additional knowledge about the expected fault.

Various stochastic filtering techniques can be used for this purpose. Kalman filtering is a powerful and common tool for this kind of applications. Using a bank of Kalman filters was pioneered by Magill (Magill, 1965) who used a parallel structure of estimators in order to estimate a sampled stochastic process. The extended Kalman filter (EKF) has become a standard technique used in a number of nonlinear systems. Because of the linearization, this filter is not applicable to highly nonlinear systems and calculating the Jacobian matrix is a drawback. To overcome this limitation, unscented Kalman filter (UKF) is developed. UKF represents a derivative-free alternative to EKF, and provides superior performance at an equivalent computational complexity. It operates on the premise that it is easier to approximate a Gaussian distribution than it is to approximate an arbitrary nonlinear function. Instead of linearizing using Jacobian matrices, UKF uses a deterministic sampling approach to capture the mean and covariance estimates with a minimal set of sample points. Although UKF is a relatively powerful tool for nonlinear system, it has some disadvantages. It is not applicable to a system with non-Gaussian noise sources; and it is not a truly global approximation based on a small set of trial points.

In recent years, a new state estimation technique, called particle filtering, has been developed (Ristic et. al., 2004; Doucet, 2001). For highly nonlinear systems with non-Gaussian additive noises, UKF is not applicable, whereas particle filters can be a good alternative (Ristic et. al., 2004).

The particle filter builds a discrete estimation of the conditional probability density given the measurement information. It has several interesting features, which make it potentially attractive for nonlinear system modeling (Goffaux and Wouwer, 2005):

- It is a nonlinear state estimation technique which does not require assumptions on the model equations.
- Discrete-time measurements are easily accommodated.
- It is a general method which handles non-Gaussian process and measurement noise.
- The particle filter is almost insensitive to state dimension and the computation load depends more on the number of particles.

The main drawback of this method is that, it is computationally expensive. However, thanks to the availability of ever-increasing computational power, these methods are already used in real-time applications in different fields (Doucet, 2001).

A particle filter approximates an unknown probability distribution using a weighted set of samples. The main objective of particle filtering is to track a variable of interest as it evolves over time. The basic idea of particle filtering is to use a random measure $\{x_k^i . w_k^i\}_{i=1}^M$ to characterize the posterior distribution of the state, $(x_k|Y_k)$, where $\{x_k^i\}_{i=1}^M$ is a set of support points (particles) with associated weights $\{w_k^i\}_{i=1}^M$, that each weight signifies quality of that specific particle. Particle filter algorithm is recursive in nature and operates in two phases: prediction and update. After each action, each particle is modified according to the existing model (prediction stage). Then, each particle's weight is re-evaluated based on the latest sensory information available (update stage). The weights are normalized such that $\sum_i w_k^i = 1$. Then the posterior density at time k can be approximated as:

$$p(x_k|Y_k) \approx \sum_{i=1}^{M} w_k^i \delta(x_k - x_k^i) \tag{13}$$

To summarize, despite the capability of IMM-based FDI method, the detection and identification of the faults in a stochastic nonlinear system remain a challenging problem. Introduction of nonlinearity and non-Gaussian noise into the system, raises new challenges in fault detection, such as how to model the nonlinearities in the systems while keeping it simple and applicable and how to handle non-Gaussian noise in the system and avoiding false alarms due to noise in the system. Furthermore, the requirements of real-time system monitoring raise additional challenges due to the constraints of online memory and computation capabilities.

In order to accomplish FDI for a nonlinear stochastic system, research is needed to overcome the following challenges:

• Develop an FDI method that can handle nonlinearities in the system by considering available computational capabilities and available online memories.

• The FDI method should be able to operate even in the presence of non-Gaussian noise in the system.

According to characteristics of particle filter presented above, it seems that particle filter can be combined with interacting multiple model (IMM) approach to accomplish fault detection and diagnosis to handle a nonlinear system with non-Gaussian noise.

Chapter 4: IMM-FDI Method with Particle Filter Estimator

4.1 IMM-PF-Based Fault Detection Method

According to pros and cons of different filters and FDI methods (provided in Table 2.1), an IMM-based FDI method based on particle filter estimation would be the most promising method to handle a nonlinear system with non-Gaussian noise. In this work a particle filter is adopted to be used as an estimator along with IMM-based FDI algorithm.

The detail of the proposed IMM-based FDI algorithm is as follows (Arulampalam et al. 2002; Cui et al., 2005; Yang et al., 2006):

Step 1: Interacting and mixing of the estimates

1.1 Compute the predicted mode probability from *k* to *k+1*, $\hat{\mu}_j(k+1|k)$:

$$\hat{\mu}_j(k+1|k) = \sum_{i=1}^{N} \pi_{ij}\mu_i(k), \quad j = 1,2,\dots,N \tag{14}$$

where N is the number of the modes.

1.2 Compute the mixing probability at *k*, $\mu_{ij}(k)$:

$$\mu_{ij}(k) = \frac{\pi_{ij} \times \mu_i(k)}{\hat{\mu}_j(k+1|k)}, \quad i.j = 1,\dots,N \tag{15}$$

1.3 Compute the mixing a priori probability density from *k* to *k+1*, $\hat{p}_{j0}(x(k)|y(k))$:

$$\hat{p}_{j0}(x(k)|y(k)) = \sum_{i=1}^{N} \hat{p}_i(x(k)|y(k)) \times \mu_{ij}, \quad j = 1, \dots, N \tag{16}$$

Step 2: Model-conditional filtering

2.1 $\forall j \in M$ draw M samples $\bar{x}_j^l(k)$ according to mixing a priori probability density $\hat{p}_{j0}(x(k)|y(k))$

2.2 Compute the predicted state from k to $k+1$, $\hat{x}_j^l(k+1)$:

$$\hat{x}_j^l(k+1) = f_j\left(\bar{x}_j^l(k), u(k)\right) + w(k), \ \ j = 1, \dots, N \ \& \ l = 1, \dots, M \tag{17}$$

where M is the number of the particles.

2.3 Compute the predicted output of samples from k to $k+1$:

$$\hat{y}_j^l(k+1) = h_j\left(\hat{x}_j^l(k+1)\right), \quad j = 1, \dots, N \ \& \ l = 1, \dots, M \tag{18}$$

2.4 Compute the probability weights at $k+1$:

$$\bar{q}_j^l(k+1) = d_{v(k,j)}\left(y(k+1) - \hat{y}_j^l(k+1)\right), \quad j = 1, \dots, N \ \& \ l = 1, \dots, M \tag{19}$$

2.5 Normalize the probability weights:

$$q_j^l(k+1) = \frac{\bar{q}_j^l(k+1)}{\sum_{l=1}^{M} \bar{q}_j^l(k+1)}, \qquad j = 1, \dots, N \ \& \ l = 1, \dots, M \tag{20}$$

2.6 Find the mean of the state over the sample set:

$$\bar{x}_j(k+1) = \sum_{l=1}^{M} q_j^l \hat{x}_j^l(k+1), \quad j = 1, \dots, N \ \& \ l = 1, \dots, M \tag{21}$$

2.7 Compute the covariance of the state over the sample set:

$$\hat{P}_j(k+1) = \sum_{l=1}^{M} q_j^l \left(\bar{x}_j(k+1) - \hat{x}_j^l(k+1)\right)\left(\bar{x}_j(k+1) - \hat{x}_j^l(k+1)\right)^T, \tag{22}$$

$j = 1, \dots, N \,\&\, l = 1, \dots, M$

2.8 Compute the probability density function for state in mode *j* after M Gaussian densities:

$$\hat{p}_j\big(x(k+1)\big|y(k+1)\big) = \sum_{l=1}^{M} q_j^l N\left(\hat{x}_j^l(k+1), \vartheta_j(k+1)\hat{P}_j(k+1)\right), \tag{23}$$

$j = 1, \dots, N \,\&\, l = 1, \dots, M$

where ϑ_j is the scaling factor and it can be calculated as $\vartheta_j = 0.5M^{-2/d_j}$ (d_j is the dimension of the state space).

Step 3: Updating the mode probability at *k+1*, $\mu(k+1)$:

3.1 Compute the residual covariance over the sample set:

$$\hat{C}_j(k+1) = \sum_{l=1}^{M} q_j^l \left(h_j\left(\bar{x}_j(k+1), u(k)\right) - \hat{y}_j^l(k+1)\right)\left(h_j\left(\bar{x}_j(k+1), u(k)\right) - \hat{y}_j^l(k+1)\right)^T \tag{24}$$

$j = 1, \dots, N \,\&\, l = 1, \dots, M$

3.2 Compute the innovations over the sample set:

$$r_j^l(k+1) = y(k+1) - \hat{y}_j^l(k+1), \quad j = 1, \dots, N \,\&\, l = 1, \dots, M \tag{25}$$

3.3 Compute the likelihood function at *k+1*, $L_j^l(k+1)$:

$$L_j^l(k+1) = N\left(r_j^l(k+1); 0, \hat{C}_j(k+1)\right), \qquad j = 1, \dots, N \ \& \ l = 1, \dots, M \tag{26}$$

3.4 Compute the mean of the likelihood over the sample set:

$$L_j(k+1) = \sum_{l=1}^{M} q_j^l L_j^l(k+1), \quad j = 1, \dots, N \ \& \ l = 1, \dots, M \tag{27}$$

3.5 Update the mode probability at *k+1*, $\mu(k+1)$:

$$\mu_j(k+1) = \frac{\mu_j(k) L_j(k+1)}{\sum_{j=1}^{N} \mu_j(k) L_j(k+1)}, \qquad j = 1, \dots, N \tag{28}$$

Step 4: Detecting fault at *k+1*:

4.1 Compute the mode probability vector at *k+1*, $\vec{\mu}(k+1)$:

$$\vec{\mu}(k+1) = [\mu_1(k+1) \quad \mu_2(k+1) \quad \dots \quad \mu_N(k+1)] \tag{29}$$

4.2 Compute the maximum value of the mode probability vector components:

$$\mu_{\max} = \max_j\{\mu_j(k+1)\}, \qquad j = 1, \dots, N \tag{30}$$

4.3 Detect the index of the maximum value of the mode probability vector components and subsequently assign: *index* = *j*

4.4 Apply fault decision-FDI logic:

If $\mu_{\max} > \mu_{Threshold}$ then a fault occurred and the index of the fault mode is *j*, otherwise no fault occurred.

The particle-filter-based approach introduced above allows statistical characterization of both discrete and continuous-valued states, as new feature data (measurements) are received. As a result, at any given instant of time, this framework provides an estimate of the probability associated with each fault mode, as well as a PDF estimate for meaningful physical variables in the system. Once this information is available within the FDI module, it is conveniently processed to generate proper fault alarms and to report on the statistical confidence of the detection routine.

Chapter 5: Anomaly Detection in Mining Haul Truck Suspension System

As mentioned earlier the objective in this project is to apply the proposed FDI method on the suspension strut problem and investigate the performance of the method. Most of the studies into FDI of suspension system apply model-based methods; and there are few studies use data-driven methods (Wei et. al., 2012). Bond graph modeling and simulation is used to solve the problem of detecting and isolating faults in vehicle suspensions (Silva et al., 2007). In (Li and Goodall, 2004; Wei et al., 2011), the authors derived a fault detection approach for the rail vehicle suspension systems based on Kalman filter. A model-based fault diagnosis scheme using an Utkin observer is presented for fault diagnosis in position sensors of a quarter standard bus suspension (Moncada and Marin, 2011). Moreover, a particles filter is applied by Li et al. (2007) for the parameter estimation and parameter changes identification to indicate the health condition of the suspension system. While it is true that a model-based condition monitoring methods can perform very well when the system model is accurate, it should be noted that they are expensive methods and hard to implement. Because of the oscillatory nature of the behavior of the suspension system and also some issues in measuring the road roughness, signal-based method is applied to this problem at the first place as the most available and simple FDI tool for this kind of systems.

5.1 Anomaly Detection in Mining Haul Truck Suspension System Applying Wavelet Analysis

A combination of continuous wavelet transform and PCA algorithm is applied for fault detection of truck suspension strut. The FDI method applied to both analytical model and operational truck data. The details of this FDI method and results are reported as a journal article (Hajizadeh and Lipsett, 2014).

To develop an automated fault detection system for a strut, some features of the anomalies should be investigated. Detectable features of anomalies are challenging to collect from a mining vehicle undercarriage. Off-high-way trucks usually have a payload measurement system to monitor production in the mine by measuring the internal pressure in the four suspension struts of each truck and applying an empirical relationship to estimate the mass of the payload. The pressure data that is gathered by these existing sensors for payload measurement can be used for suspension strut fault detection, making implementing an FDI system much more simple than one requiring installation of additional sensors and a data logger. Condition monitoring of the strut may thus also improve payload monitoring accuracy.

Pressure of the fluid within the strut is related to a payload weight within the dump body and also it is subject to oscillations while the dump is traveling. Although these oscillations will vary in frequency and magnitude, their natural frequencies and their frequency components with relative maximum energy are almost the same in a normal strut. Therefore local faults and transient phenomena in dynamic systems can be identified from its effect in the frequency components of the signal.

Pressure signals have transient feature components that have local energy distribution in the time and frequency. A standard power spectrum analysis is therefore likely to yield misleading results. Wavelet analysis is one of the most efficient methods to analyze oscillatory signals for time-varying systems (Lin and

Zuo, 2003) and it can be used instead of standard power spectrum analysis to tackle time varying systems. The continuous wavelet transform (CWT) is a powerful tool to approach the natural frequencies of the system and the frequency components of the signal with relative maximum energy, and can be used on sequences of sampled data.

Once the frequency features have been extracted from strut pressure signals by wavelet, a ranking and classification scheme is applied to determine whether a fault is likely to be present or not. Principal Component Analysis (PCA) is one approach for feature ranking and classification based on the assumption of linear contributions of features to the classification.

5.1.1 Fault Detection and Identification with CWT and PCA

In this study continuous wavelet transform (CWT) is considered because it reveals the signal content in far greater detail than discrete wavelet transform (DWT) (Özgönenel et al., 2005).

The continuous wavelet transform is defined by the following equation (Lee and White, 2000):

$$W_x^{\psi}(a,b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}^{*}(t)dt = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\psi^{*}\left(\frac{t-b}{a}\right)dt \qquad (31)$$

The transformed signal is a function of two variables, a and b, the translation and scale parameters, respectively, $\psi(t)$ is the transforming function localized both in time and frequency, called the mother wavelet, a template for generating the other window functions. The function $\psi_{a,b}(t)$ is obtained by applying the operation of shifting in the time domain (b-translation) and scaling in the frequency domain (a-translation) to copies of the mother wavelet.

Scale b is the position parameter in wavelet transform and it determines the position of the wavelet function on time axes. Scale a is the scale parameter in wavelet transform and it determines how much the wavelet function is stretched or compressed. There is an inverse relationship between scale a and frequency of the signal. The higher scales correspond to the most "stretched" wavelets that can detect slowly changing features or low frequencies. The lower scales correspond to compressed wavelets that can detect finer features with higher frequencies. It should be noted that there is no precise relationship and mapping between the scale and the frequency. They generally can be related by defining a pseudo-frequency, F_a, corresponding to scale a. The center frequency, F_c, of the wavelet can be used to calculate the corresponding frequency of scale a:

$$F_a = \frac{F_c}{a\mathrm{T}} \tag{32}$$

where a is scale, T is the sampling period, F_c is the center frequency of the wavelet function in Hz and F_a is the pseudo-frequency corresponding to the scale a, in Hz.

Any signal processing performed on a computer must be performed on a discrete signal and CWT in computer is actually a discrete process. However, unlike the discrete wavelet transform, the CWT is not limited in the set of scales and positions at which it operates and it can operate at every scale. The continuous nature of the wavelet function is kept up to the point of sampling the scale-translation grid used to represent the wavelet transform. The CWT is also shifted smoothly over the full domain of the analyzed function and it can be considered continuous in terms of shifting. It is much easier to interpret the result obtained from CWT and to draw conclusion from the data.

Various wavelets are available for wavelet applications and wavelets that have a good time-frequency localization property are more desirable to use for fault detection (Zhao et al., 2004). In this work, different analyzing wavelets, are examined (Haar, Mexican hat, Daubechies and complex Morlet) and among them Morlet wavelet shows better performance in detecting fault.

The Morlet wavelet has large similarity with the impulse generated by strut faults; and so it is applicable to extract features of the fault out of the signal (Lin and Qu, 2000).

The definition of complex Morlet wavelet function is:

$$cmor(x) = \frac{1}{\sqrt{\pi F_b}} e^{2i\pi F_c x} e^{-x^2/F_b} \tag{33}$$

First both real and imaginary part of the Morlet wavelet is considered and the result shows that real part is sufficient to extract the features from pressure data set. Therefore, in this study, the real-part of the complex Morlet wavelet is used:

$$\psi(t) = \left(\frac{\beta}{\sqrt{2\pi}} \cdot e^{-\beta^2 t^2/2}\right) \cdot cos(\omega_c t) \tag{34}$$

The square of the modulus of the CWT can be interpreted as an energy-density distribution over the *(a,b)* time-scale plane.

The energy of a signal on this plane is mainly concentrated around the so-scaled ridges of the wavelet transform. The contribution of signal energy in the specific scale *a*, and the translation *b*, is defined by a two-dimensional wavelet energy-density function *E(a,b)* called a scalogram:

$$E(a,b) = \left|W_x^{\psi}(a,b)\right|^2 \tag{35}$$

Wavelet transform coefficients can enhance measured data from system and create indicators which are useful for detection of faults by making the fault more apparent in the time-frequency energy distribution (scalogram). To have an operational fault detection system, a classifier is needed to automate the classification based on features extracted by the wavelet. In this study, PCA is used for classification of these features.

PCA is a multivariate analysis technique that is used for dimension reduction. It projects a high-dimensional space onto a space that has significantly fewer dimensions and it simplifies the feature selection by reducing the feature space dimension (Chiang et al., 2001). PCA can be defined as a linear transformation of the original correlated data into a new set of uncorrelated data. In this way, PCA is a good technique to transform the set of original wavelet coefficients into a new set of uncorrelated coefficients that explain the trend of the measured data. A scalogram obtained from a wavelet has a large space of scales; and PCA can project and reduce this space onto a smaller space. The indicators obtained from scalogram can be stored in a matrix which entry (i,j) of this matrix is the value of the scalogram for scale a_j computed at time t_i. The scalogram is examined as a function of the scales and only those scales that show more sensitivity are retained to form the reduced subspace.

Consider $X \epsilon \mathcal{R}^{n\times m}$ to be the matrix of scalogram after normalizing it to have zero mean and unit variance, where each row corresponds to a time instant and each column is associated to one of the selected scales. PCA transforms X by combining the variables as a linear weighted sum as:

$$T = XV \tag{36}$$

where V is the principal component loadings matrix, and T is the principal component score matrix which contains the principal components of the system. The principal component loadings denote the direction of the hyperplane that captures the maximum possible residual variance in the measured variables, while maintaining orthonormality with the other loading vectors.

V can be obtained by constructing the covariance matrix of X and applying singular value decomposition:

$$S = \frac{1}{n-1} X^T X \tag{37}$$

$$S = V \Lambda V^T \tag{38}$$

where the diagonal matrix $\Lambda \epsilon \mathcal{R}^{m \times m}$ contains the non-negative real eigenvalues of decreasing magnitude ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$). $V \epsilon \mathcal{R}^{\mathrm{m \times m}}$ is a unitary matrix, and the columns of matrix V are the eigenvectors of S (also called loading vectors). The eigenvectors define the hyperplane that captures the maximum possible variance in the X and the eigenvalues indicate the variance captured by the corresponding eigenvector.

The transformation matrix $P \epsilon \mathcal{R}^{m \times a}$ is generated choosing the first a eigenvectors or a first columns of V corresponding to the first a principal eigenvalues. The purpose of P is to reduce the dimension space of the measured variables. The a largest singular values are typically retained to capture the variations of the data optimally while minimizing the effect of random noise. The projections of the observations in X into the lower-dimensional space are contained in the score matrix $\hat{T}$ (Chiang et al., 2001):

$$\hat{T} = XP \tag{39}$$

The projection of $\hat{T}$ back into the m–dimensional observation space is $\hat{X}$

$$\hat{X} = \hat{T} P^T \tag{40}$$

where the difference between X and $\hat{X}$ is the residual matrix E:

$$E = X - \hat{X} \tag{41}$$

The PCA model partitions the measurement space into two orthogonal subspaces $\hat{X}$ and E which are called the score space and residual space, respectively. The score space represents data variations according to the principal component model and the residual space represents data variation not captured by the model.

Having established a PCA model based on the scalogram of healthy strut pressure, Hotelling's T^2 statistic and square prediction error (SPE) or Q statistic can be used to do the monitoring. The T^2 and Q statistics, along with their appropriate thresholds, are able to detect different types of faults, and they can be used together to utilize their abilities (García-Álvarez, 2009). The T^2 and Q statistics are common PCA metrics that indicate how well a particular sample fits a specific PCA model (Wise and Gallagher, 1996). The Q statistic refers to the sum of the squared residuals. Figure 5.1 shows that the Q statistic is the distance that a data point is outside the subspace of the PCA model. The T^2 statistic, on the other hand, is the distance from the origin of the PCA model to the data point, within the subspace of the PCA model.

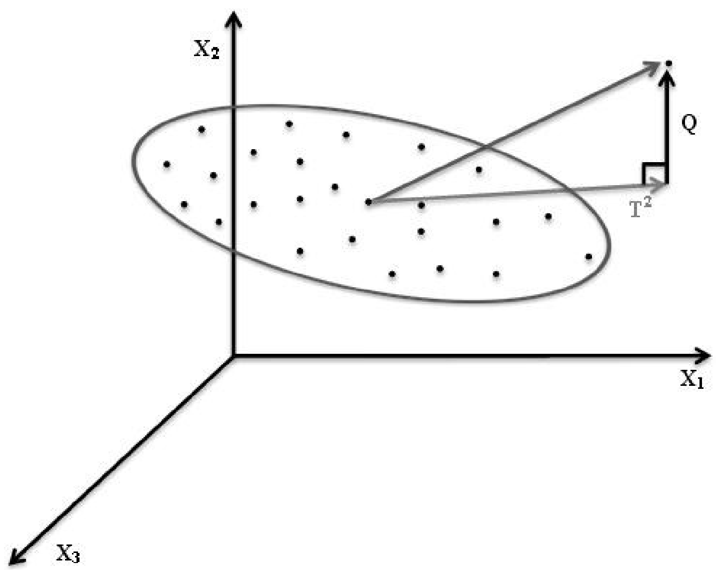

Figure 5.1: The $\mathbf{T^2}$ and $\boldsymbol{Q}$ statistics representation for a data point with respect to PCA model

Hotelling's T^2 statistic can be calculated as the sum of squares of a new process data vector $\hat{x}$ characterized by the normal operation:

$$T^2 = \hat{x}^T P \Lambda_a^{-1} P^T \hat{x} \tag{42}$$

where Λ_a is a square matrix formed by the first a rows and columns of Λ.

The system can be monitored at each sampling time by comparing the Hotelling's T^2 statistic obtained for each data with a predefined threshold. This threshold can be computed as follows:

$$T^2 \leq T_\propto^2 = \frac{(n^2-1)a}{n(n-a)} F_\alpha(a, n-a) \tag{43}$$

where $F_\propto(a, n-a)$ is the upper $100\alpha\%$ critical point of the F-distribution with a and *n-a* degrees of freedom. The T^2 statistic with Equation (43) defines the normal process behavior, and an observation vector outside this region indicates a fault in the system. A T^2 statistic can be interpreted as measuring the systematic variations of the process and violation of the T^2 statistic indicates that variations are out-of-control. The T^2 statistic is based on the first a principal components, therefore it is sensitive to inaccuracies in the PCA space corresponding to the smaller singular values.

The portion of the observation space corresponding to the $m - a$ smallest singular values can be monitored more robustly by using the Q statistic. The Q statistic is calculated as the sum of squares of the residuals; and it can detect new events and random variations of the system. The scalar value Q is associated with the noise in the system and it shows fitness of the sample to the PCA model:

$$Q = e^T e \tag{44}$$

$$e = (I - PP^T)x \tag{45}$$

where e is the residual vector. Another threshold Q_a for Q can be defined as follows:

$$Q_a = \theta_1[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2}]^{\frac{1}{h_0}} \tag{46}$$

with

$$\theta_i = \sum_{j=a+1}^{m} \sigma_j^{2i} \qquad h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \tag{47}$$

where c_α is the value of the normal deviate corresponding to the $(1-\alpha)$ percentile. Given a level of significance, α, violation of the Q statistics from threshold Q_a indicates that the characteristic of the measurement noise has changed and or an unusual event has occurred. Every time an unusual event happens, it changes the covariance structure of the model and therefore the residual passes the threshold.

5.1.2 Half Car Simulation for Verification of the FDI Method

To study the performance of the FDI method for the strut application in a controlled situation with different road and strut conditions, an analytical half-car dynamic suspension system is considered. This analytical model allows normal and faulty strut behaviour to be modeled; and their response to the same road pattern (system input) can be analyzed to classify normal and faulty struts. This model of strut can be utilized in application of IMM-based FDI method to analytical strut model as well. A general half-car model is used in this analysis as shown in Figure 5.2.

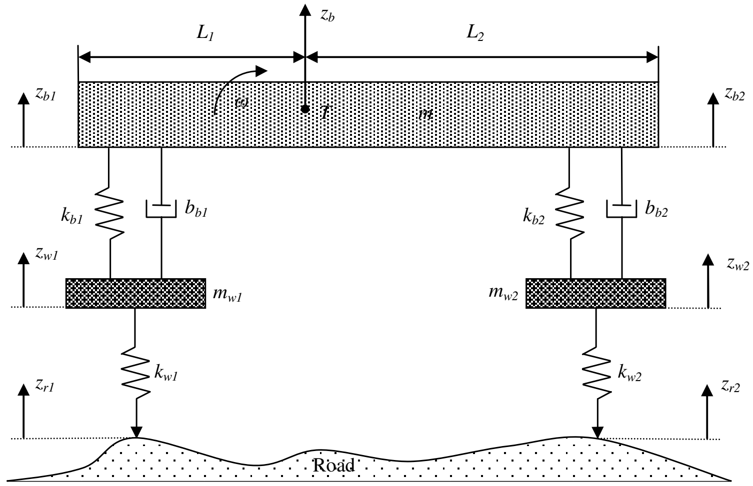

Figure 5.2: One-half-car model (Modified from Stribrsky et al., 2003)

The vehicle suspension model consists of a single sprung mass (car body) connected to two masses representing the front and rear wheel assembly masses at each end. The sprung mass is free to heave and pitch, which comprises four states: vertical displacement, vertical velocity, pitch angular displacement, and pitch angular velocity. The other masses are free to bounce vertically with respect to the sprung mass. The suspension elements between the sprung mass and wheel/axle masses are initially modeled as nonlinear viscous dampers and linear spring elements, while the tires are modeled as simple linear springs without damping components. This model is used to simulate the response of the suspension system of a vehicle when driven over a rough terrain.

The equations of motion for the car body and the front and rear wheels are (Stribrsky et al., 2003):

$$m\ddot{z}_b + b_{b1}(\dot{z}_{b1} - \dot{z}_{w1})|\dot{z}_{b1} - \dot{z}_{w1}| + b_{b2}(\dot{z}_{b2} - \dot{z}_{w2})|\dot{z}_{b2} - \dot{z}_{w2}| + k_{b1}(z_{b1} - z_{w1}) + k_{b2}(z_{b2} - z_{w2}) = 0 \quad (48)$$

$$I\ddot{\omega} + L_1\big(b_{b1}(\dot{z}_{b1} - \dot{z}_{w1})|\dot{z}_{b1} - \dot{z}_{w1}| + k_{b1}(z_{b1} - z_{w1})\big) - L_2\big(b_{b2}(\dot{z}_{b2} - \dot{z}_{w2})|\dot{z}_{b2} - \dot{z}_{w2}| + k_{b2}(z_{b2} - z_{w2})\big) = 0 \tag{49}$$

$$m_{w1}\ddot{z}_{w1} - b_{b1}(\dot{z}_{b1} - \dot{z}_{w1})|\dot{z}_{b1} - \dot{z}_{w1}| - k_{b1}(z_{b1} - z_{w1}) + k_{w1}(z_{w1} - z_{r1}) = 0 \tag{50}$$

$$m_{w2}\ddot{z}_{w2} - b_{b2}(\dot{z}_{b2} - \dot{z}_{w2})|\dot{z}_{b2} - \dot{z}_{w2}| - k_{b2}(z_{b2} - z_{w2}) + k_{w2}(z_{w2} - z_{r2}) = 0 \tag{51}$$

Table 5.1 summarizes the parameters used to simulate the vehicle in a half-car model (Gao et al., 2007).

Table 5.1: Parameters of the half car model (Gao et al., 2007)

Parameters	**Values**	**Parameters**	**Values**
m	1749 *kg*	k_{b1}	66824 *N/m*
I	3443 $kg.m^2$	k_{b2}	18615 *N/m*
m_{w1}	87 *kg*	b_{b1}	7190 *N.s/m*
m_{w2}	140 *kg*	b_{b2}	7000 *N.s/m*
L_1	1.271 *m*	k_{w1}	101115 *N/m*
L_2	1.716 *m*	k_{w2}	101115 *N/m*

The relative displacement $z_{b1} - z_{w1}$ and velocity $\dot{z}_{b1} - \dot{z}_{w1}$ of each tire with respect to vehicle chassis, respectively, can be calculated by solving the above system in the first-order form.

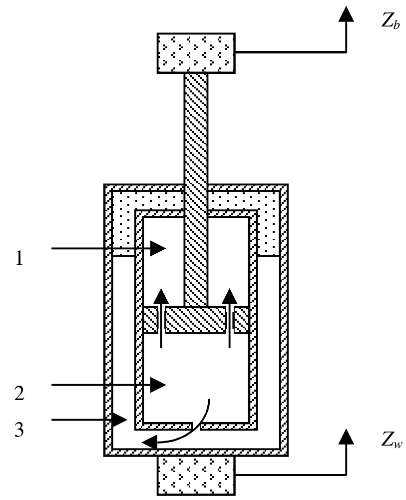

Figure 5.3: Schematic of a hydraulic damper strut

A schematic of a strut assembly is illustrated in Figure 5.3, which consists of a piston that moves up and down in fluid-filled cylinder. The piston is fastened via the piston rod to the chassis of the vehicle and the cylinder is connected to the wheel. The cylinder is divided in two parts by the piston. The interior volume above the piston is the rebound chamber, and the volume below the piston is the compression chamber. A chamber is surrounding the cylinder namely reserve chamber and it is partially filled with shock absorber fluid (liquid) and partially filled with a gas. During the compression stroke, fluid from the compression chamber is forced up through the piston orifices and valves into the rebound chamber. Fluid equivalent to rod volume being inserted into the rebound chamber is also forced down through the orifices at the bottom of the cylinder into the reserve chamber. During a rebound stroke, fluid returns back to compression chamber.

Here, for simplification of the model, fluid in the strut is considered to be incompressible because the change in the density of oil in the strut due to pressure change in strut is negligible and also considering the pressure inside the strut, it is assumed that there is no change in the volume of cylinder due to change in fluid pressure. Also, all the notches and control valves in the strut are modeled as a simple orifice.

Orifice equations are used to model flow. To satisfy continuity requirements, net flow from the compression chamber must be equal to the equivalent fluid displacement due to piston movement. It can be shown in the following equation (Harvey et al., 1977):

$$Q_c = Q_{21} + Q_{23} = \left(A_p + A_r\right)\dot{z} \tag{52}$$

where Q_{21} is flow from compression chamber into rebound chamber, Q_{23} is flow from compression chamber into reserve chamber, and $\mathrm{A_{p}}$, and $\mathrm{A_r}$ are the area of the piston in the rebound side and the area of the piston rod, respectively. $\dot{z}$ is relative velocity of tire with respect to chassis i.e., $(\dot{z}_{b1} - \dot{z}_{w1})$. The net flow rate to or from rebound chamber is equal to the equivalent flow rate due to the piston movement minus volume of the rod inserting into or withdrawing from cylinder. This is equal to flow through the orifices:

$$Q_{21} = A_p\dot{z} = -nC_{d1}a_1\sqrt{2\,|p_{21}|/\rho}\,\mathrm{sgn}(p_{21}) \tag{53}$$

Similarly, the oil flow through the cylinder orifice Q_{23} can be expressed as

$$Q_{23} = A_r\dot{z} = -C_{d2}a_2\sqrt{2\,|p_{23}|/\rho}\,\mathrm{sgn}(p_{23}) \tag{54}$$

From Equations (53) and (54), the pressure difference across the compression chamber and rebound chamber and also pressure difference across compression chamber and reservoir are thus expressed as:

$$p_{21} = -\frac{\rho}{2n^2{C_{d1}}^2}\left(\frac{A_p}{a_1}\right)^2 |\dot{z}|\dot{z} \tag{55}$$

$$p_{23} = -\frac{\rho}{2{C_{d2}}^2}\left(\frac{A_r}{a_2}\right)^2 |\dot{z}|\dot{z} \tag{56}$$

Assuming that the gas in the reservoir is an ideal gas that undergoes a reversible, adiabatic process, then the instantaneous pressure of the gas column in the reservoir $p_3$is given by:

$$p_3 = p_0(v_0/v_3)^\gamma \tag{57}$$

where p_0, v_0 are the initial pressure and volume of the gas in the reservoir (assumed to be constant for given mass). The instantaneous volume of the gas column v_3 is related to the relative motion across the hydraulic damper:

$$v_3 = v_0 + A_r z \tag{58}$$

This formulation assumes nonlinear constitutive relationships between damper pressures and vertical velocity of tire with respect to vehicle chassis. From initial values of p_0 and v_0, Equations (55) to (58) can be solved to find $p_{21}$and p_2, pressure difference across compression chamber and rebound chamber and pressure in the compression chamber, respectively.

Table 5.2 shows the parameters in the damper equations that are being used in this study (Dixon, 2007; Harvey et al., 1977).

Table 5.2: Parameters of the hydraulic damper

Parameters	Values	Parameters	Values
ρ	$860\,kg/m^3$	C_{d1}	0.7
C_{d2}	0.1	A_p	$2.1\times10^{-3}m^2$
A_r	$2\times10^{-3}m^2$	a_1	$4\times10^{-5}m^2$
a_2	$4\times10^{-5}m^2$	p_0	$334\,kPa$
v_0	$1.2\times10^{-3}m^3$	γ	1.4

Utilizing the defined half car model and constitutive relationship for strut, a propagation simulation will yield pressure in the strut by using known road condition data as an input.

The three most likely faults in a strut are:

1. Under-pressure strut; which occurs because of leakage of fluid from the strut, and can decrease the damping effect of strut and its efficiency;

2. Leaking through the clearance between the piston and cylinder inside the strut;

3. Broken control valve in the strut. There are control valves or blow-off valves inside the strut. They have a pre-loaded valve spring and will not open until a specified differential has built up across the valve. These valves may be stuck or their spring may not work properly. Thus, flow through the valves will change and as a result damping effect of strut may change.

To study the performance of FDI method, the most likely faults in strut need to be simulated separately in the lumped-parameter suspension model. Under-pressure strut is modeled by considering the initial pressure of the strut less than the required amount. Leakage from piston into cylinder can be analyzed by considering it as flow between parallel plates when one plate moves in its own plane and adding this term into the equation that shows flow from compression chamber into the rebound chamber. In the strut model that we have developed for simplification, control valves on piston are modeled by two simple orifices. In healthy model of strut two orifices are considered on the piston. To simply model third fault or more precisely broken control valve in the strut, just one orifice is considered on the piston to simulate that one of the control valves is stuck.

Road profile is the main source of excitation in simulation of the vehicle dynamic and an accurate road model is very important. Each road profile is different from others and even for analyzing different parts of the same road they can be considered to be a sample of a random population (Morello et al., 2011). A road profile can fit the category of stationary Gaussian random noise and a pseudo-random input can approximate fairly well a real road profile (Dodds and Robson, 1973; Rill, 2012). So, in this study irregularities of a road are described by a zero-mean Gaussian random noise produced by MATALB. It can model usual irregularities of different shapes in the road but it cannot model the sharp variations of the road profile such as potholes. Furthermore, Off-road haul trucks are usually driven on rough, unpaved roads and road profile for this specific application might be different and irregularities might have different pattern and there might be some sharp variations in the road profile that should be modeled differently. This needs more investigation on the road profile of the mining benches; but for sake of simplicity in this part for data-driven FDI method Gaussian irregularities are considered, even though system inputs may in fact be non-Gaussian in some operating conditions. To study the effect of different faults in the strut on the spectral representation of vertical motion of the vehicle body and on the strut pressure, four different vehicle models (one healthy and three different faulty models) are simulate to go over a road which is modeled as a zero-mean Gaussian random noise with a constant speed in MATLAB. The first scenario is a reference case when strut is operating in its normal condition and it has no fault. Three other scenarios were examined, representing three fault cases. Solving the differential equations in MATLAB using an ODE solver with fixed time step yields the response of the system over a period of time: displacement and rotation of chassis, displacement of wheels, and pressure in the dampers. In this study pressure in the damper (the pressure difference between the compression chamber and rebound chamber) is of particular interest.

Wavelet analysis is applied to the damper pressure data obtained from simulation to extract time-frequency features and to assess whether fault detection and identification can be done using this technique, for different kind of faults.

Sometimes it is easier to compare wavelet transforms of two different signals by calculating the total signal energy distribution at a specific scale *a* for each signal. Total signal energy distribution at a specific scale can be calculated by adding together all signal energy in different translations at that scale, as follows:

$$E(a) = \int_{-\infty}^{+\infty} E(a,b)db \qquad (59)$$

Figures 5.4 and 5.5 show energy diagram results of pressure in the strut, obtained by application of the wavelet transform to four different normal and faulty scenarios to determine the dominant frequency. Figure 5.4 is the total signal energy distribution for different scenarios to different pseudo-random road profiles. Figure 5.5 is the total signal energy distribution for different scenarios to the same pseudo-random road profile. By comparing these figures, generally it is shown that the increase in energy is a result of fault one and two in the system. Moreover, changes of the peak frequency can be seen in faults one and three. These results indicate that this approach can be applied to detect a faulty suspension strut from normal suspension behaviour.

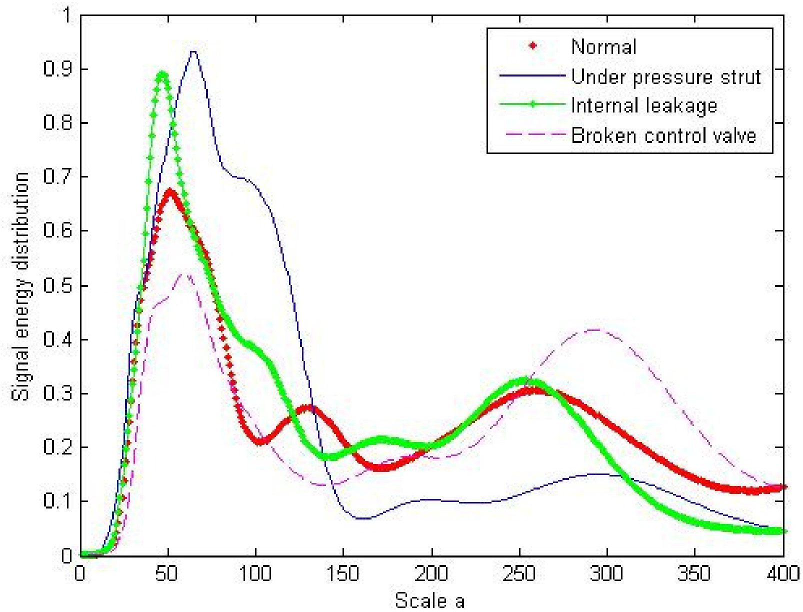

Figure 5.4: Simulated signal energy distribution of pressure signal p_2 at a specific scale a for different road profiles

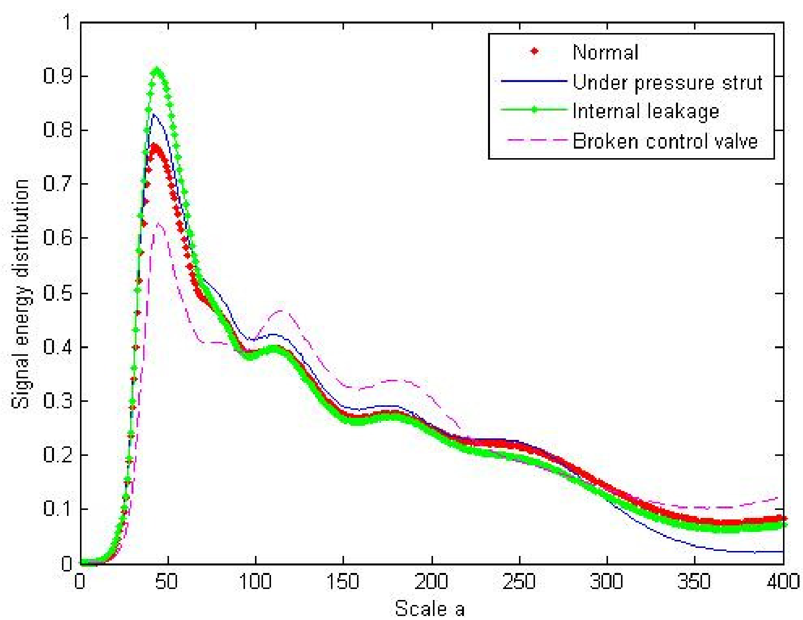

Figure 5.5: Simulated signal energy distribution of pressure signal p_2 at a specific scale a for the same road profile

For the industrial case, there is a potential issue with sampling rate. The sampling rate in the data logger installed on a haul truck is very slow, only 1 Hz. To study the effect of down-sampling on the performance of the CWT fault detection method, the pressure signal from the simulation of the half-car model was down-sampled by a factor of six. CWT is then applied to this down-sampled data. The result is represented in Figure 5.6. Although down-sampling changed the peak frequency of the signals, different faults still have different energy of the signals or the peak frequencies.

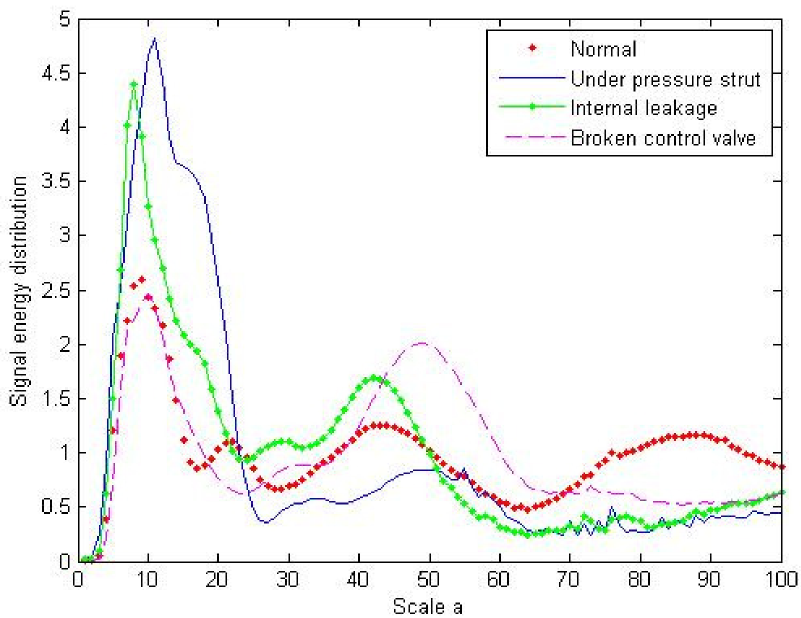

Figure 5.6: Simulated signal energy distribution of down-sampled data pressure signal p_2 at a specific scale a for different road profile

5.1.3 Application of the FDI Method on Actual Truck Data

To investigate effectiveness and performance of the proposed FDI method in an industrial application, the method is applied to six-month operating strut pressure time-series data of a haul truck recorded at 1 Hz sampling rate. This six-month dataset contains normal and faulty strut pressure data, as well as a maintenance history of when struts were fixed in this period. The analysis is based on the assumption that strut failure artifacts could be found from a process data set and the maintenance history, using continuous wavelet transform.

By knowing time and date that struts were fixed, two sets of healthy and one set of faulty pressure data can be separated from the overall set of data for the six-month period. One of the healthy data set is used to train the PCA model and two other data sets are used to test it.

First, CWT is applied to training healthy strut pressure data and resulted scalogram is shown in Figure 5.7. The energy distribution obtained from applying CWT to healthy pressure data is used to train the PCA model. Number of principal components as well as upper control limits for T^2 and Q statistics can be obtained from this healthy pressure data set.

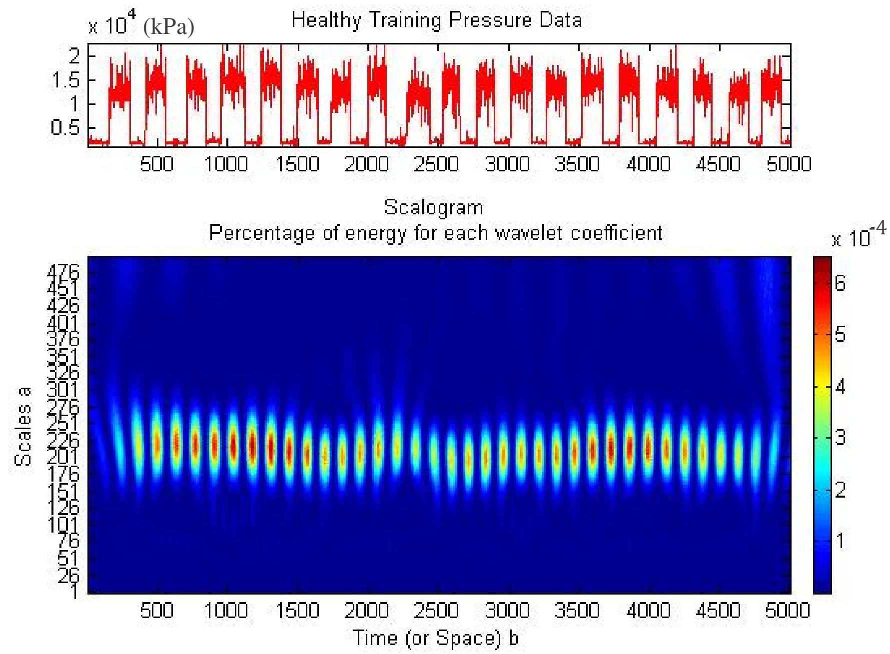

Figure 5.7: Scalogram for training healthy strut pressure data set

Once the PCA model is formed, the next task is to test performance and reliability of this model by applying this to a set of faulty pressure data set as well as healthy pressure data set. Therefore, after training the PCA model by healthy training pressure data as mentioned above another set of healthy data and a set of faulty data are separated from the rest of pressure dataset. CWT is applied to these datasets and energy distribution for each time and frequency are obtained. The resulting scalograms of these datasets are shown in Figures 5.8 and 5.9.

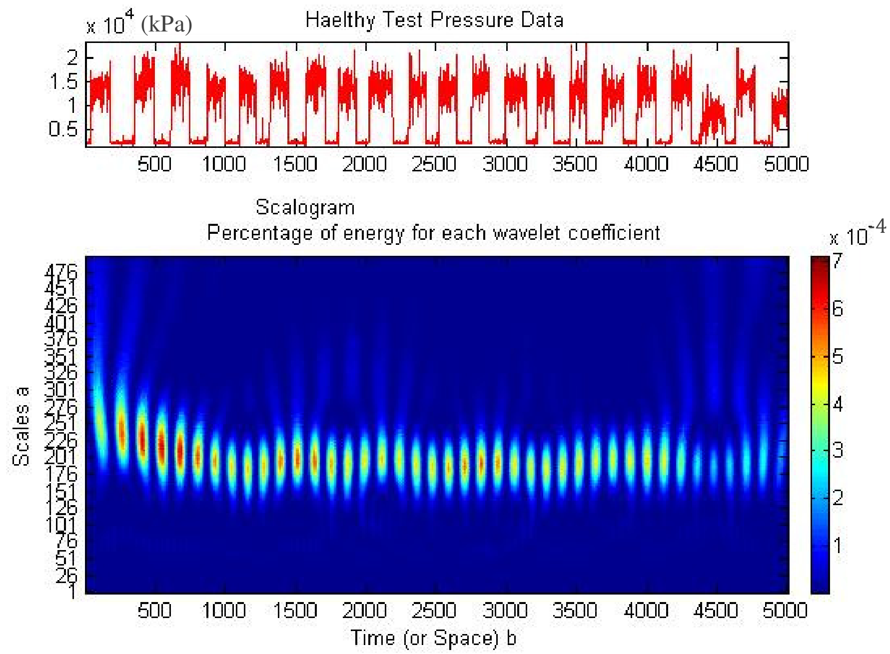

Figure 5.8: Scalogram for test healthy strut pressure data set

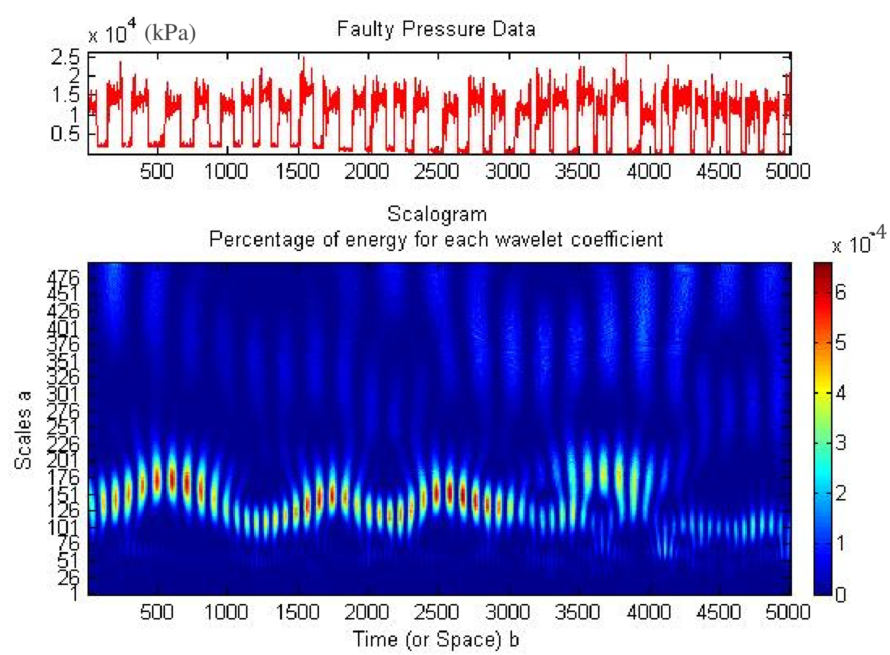

Figure 5.9: Scalogram for faulty strut pressure data set

CWT processing yields two matrices, which are then scaled using the mean and standard deviation of the training data. The PCA model obtained from training data set is used to evaluate the $T^2$and Q statistics for the test set of healthy and faulty data. If either of the T^2 or Q statistics test exceeds the upper limit, then this measurement is considered to be an alarm for a binary classification of no-fault vs. fault. If there is some consecutive established number of alarms, according to some criterion to reject excursions due to noise for a single datum, then an uncommon event has occurred. The result of the T^2 and Q statistics for healthy and faulty data set and corresponding threshold for each test are represented in Figure 5.10 and 5.11, respectively. In Figure 5.10, monitoring results for healthy data are shown. It can be seen from the figures that the value of T^2 and Q statistics tests exceeds the thresholds for some sample numbers; but tests were passed and the thresholds are set to be greater than value of T^2 and Q statistics for most of the test. In Figure 5.11, a faulty strut is monitored using PCA. In this case, both the $T^2$and Q statistics are greater than the predefined threshold, from which the FDI method determines that there is a fault in the strut.

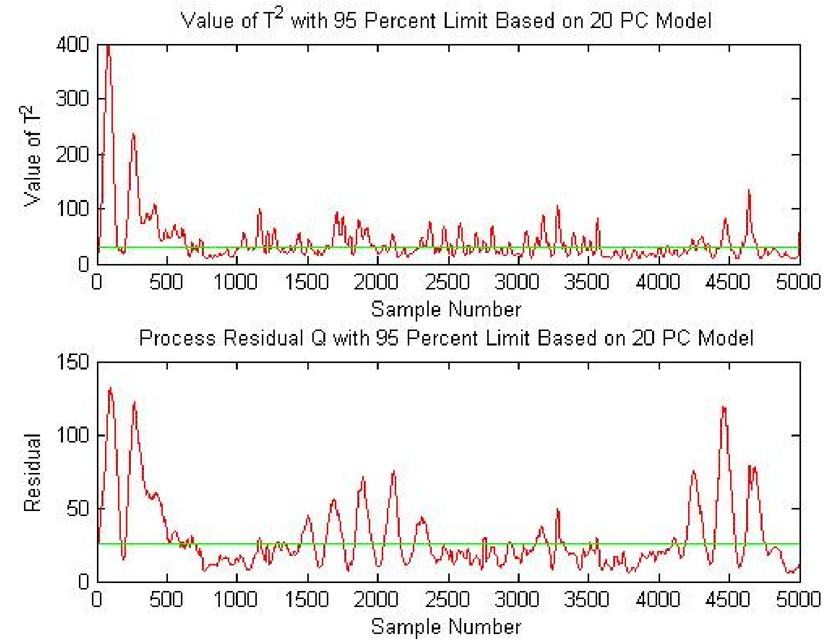

Figure 5.10: PCA result for healthy strut pressure data set

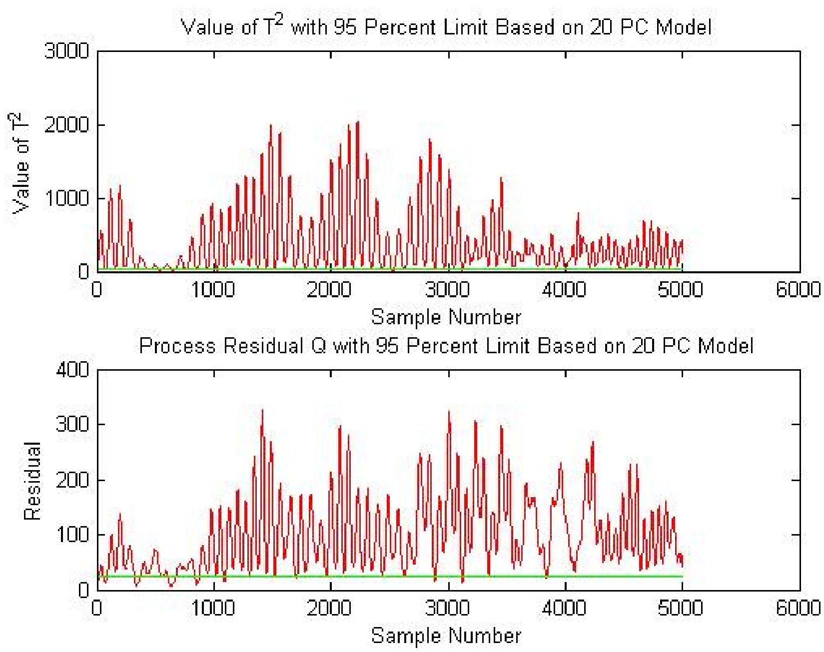

Figure 5.11: PCA result for faulty strut pressure data set

Although vibration-based condition monitoring methods are well developed, the linear nature of the eigenvalue decomposition is likely sensitive to changes in the system inputs (different road patterns at different times). Data-driven method used in this study is a good and simple solution to problem of fault detection in strut, but as it is observable from Figure 5.10 there are considerable number of false alarms in the results. The nonlinear nature of the process and also non-Gaussian noises in the system leads to poor results in applied data-driven FDI approach. Also, as mentioned above, the sampling rate in the data logger installed on haul truck is very slow, only 1 Hz, and this low sampling rate might be lower than the Nyquist frequency of the system and can cause aliasing in the signal lead to losing some information in the measured signal and inaccurate fault detection. Road roughness in paved road can be roughly modeled as white noise with Gaussian distribution. However, Variations of nonstationary effects such as root mean square of the road profile are bigger in unpaved mine benches and also transient changes are more probable compared to a paved road. Therefore, it is not realistic to consider road profile in mining benches as a signal with Gaussian distribution. For this reason, it is more general to model road pattern in mine benches as a system input with non-Gaussian distribution. This non-Gaussian input and environmental noise sources and also nonlinearities in the system can lead to non-Gaussian process and measurement noises (Yin and Zhang, 2012). Therefore, there is a need for more complicated method to tackle these complexities in the system. On the other hand, data-driven method is not able to accomplish both fault detection and identification and a fault diagnosing scheme needs to be applied to attain this task. The proposed model-based method in section 4.1 can handle these deficiencies in applied data-driven method and it can be applied on nonlinear systems with non-Gaussian noises and also it has the advantage of being able to perform both fault detection and fault diagnosis. In the coming sections of this thesis a hybrid model-based FDI method is applied primarily on a bench mark problem to show its performance in a more controlled situation and then on a suspension strut as the main focus in this research.

5.2 Application of IMM-Based FDI Method in a Two Tank Hydraulic System as a Benchmark Problem

The proposed FDI scheme has been implemented on a two-tank system to verify its performance on a common benchmark. The motivation for this case study was to determine the relative advantages of applying particle filter in IMM structure for fault detection of nonlinear systems with non-Gaussian noise compare to application of Kalman filter. Due to nonlinear behavior of a two-tank hydraulic system, it can be used as a good example to show the utility of nonlinear fault detection and isolation (FDI) techniques and it has been widely used as a benchmark study in many contributions (Khan et al., 2010). This benchmark problem is a hydraulic system and it can give some insight into modeling and fault detection of the suspension strut problem which can be considered as a hydraulic system as well.

The system under consideration comprises of two identical cylindrical tanks with a cross section S. The schematic diagram of the considered system is shown in Figure 5.12. The tanks are coupled by a connecting cylindrical pipe with a cross section S_n and an outflow coefficient μ_{12}. The nominal outflow is located at tank 2; it also has a circular cross section S_n and an outflow coefficient μ_{20}.

Using the mass balance equations, the system can be described by the following equations:

$$S\frac{dl_1(t)}{dt} = q_1(t) - q_{12}(t) \tag{60}$$

$$S\frac{dl_2(t)}{dt} = q_{12}(t) - q_{20}(t) \tag{61}$$

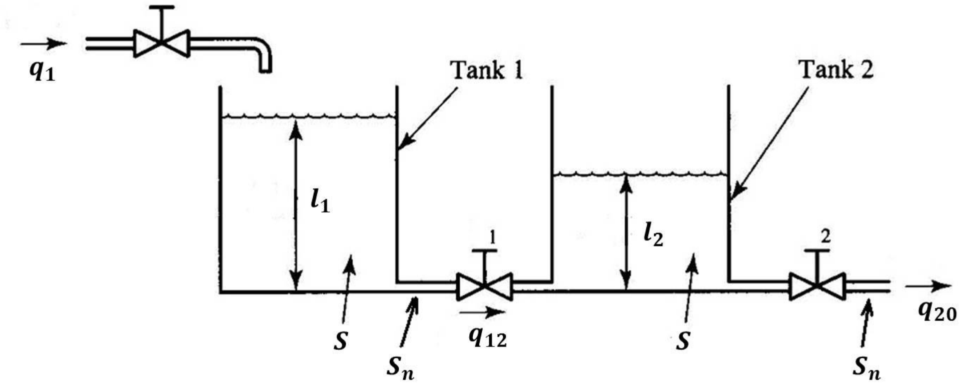

Figure 5.12: Two-tank system

Where l_1 represents level in tank 1, l_2 represents level in tank 2, q_1 represents the inflow rate to tank 1, q_{12} represents the flow-rate from tank 1 to tank 2, and q_{20} represents the outflow rate from tank 2. These unmeasured flow rates can be determined using the Torricelli-rule as

$$q_{12}(t) = \mu_{12} S_n \text{sgn}[l_1(t) - l_2(t)] \times \sqrt{2g|l_1(t) - l_2(t)|} \tag{62}$$

$$q_{20}(t) = \mu_{20} S_n \sqrt{2g l_2(t)} \tag{63}$$

The system model is described by two nonlinear first-order differential equations as:

$$\frac{dl_1(t)}{dt} = \frac{q_1(t)}{S} - \frac{\mu_{12} S_n \text{sgn}[l_1(t) - l_2(t)] \times \sqrt{2g|l_1(t) - l_2(t)|}}{S} \tag{64}$$

$$\frac{dl_2(t)}{dt} = \frac{\mu_{12} S_n \text{sgn}[l_1(t) - l_2(t)] \times \sqrt{2g|l_1(t) - l_2(t)|}}{S} - \frac{\mu_{20} S_n \sqrt{2g l_2(t)}}{S} \tag{65}$$

The numerical values of the plant parameters are listed in Table 5.3.

Table 5.3: Parameters of the two-tank system (Theilliol et al., 2002)

Variable	Symbol	Value
Tank cross sectional area	S	$1.54 \times 10^{-2}\mathrm{m}^2$
Inter tank cross sectional area	S_n	$5 \times 10^{-5}\mathrm{m}^2$
Outflow coefficient	μ_{12}	0.46
	μ_{20}	0.6
Maximum flow rate	q_1	$1 \times 10^{-4}\mathrm{m}^3\mathrm{s}^{-1}$

Two different faults are considered in this system. These faults are as follows:

- Leak in tank 1: The leak is assumed to be circular in shape and with Outflow coefficient μ_{12}. So, the leakage rate is $\mu_{12}S_n\sqrt{2gl_1(t)}$
- Leak in tank 2: Analogously to the case of leakage in tank 1, the leakage rate is $\mu_{20}S_n\sqrt{2gl_2(t)}$

To apply IMM-based fault detection method on this case study three models are needed: one model to simulate healthy mode and two more models to simulate faulty modes of the systems. Leak in tank 1 is modeled by adding another term to Equation (64) that models the leakage from tank 1. Models for two-tank system with leakage in tank 1 are presented as follows:

$$\frac{dl_1(t)}{dt} = \frac{q_1(t)}{S} - \frac{\mu_{12}S_n\mathrm{sgn}[l_1(t)-l_2(t)]\sqrt{2g|l_1(t)-l_2(t)|}}{S} - \frac{\mu_{12}S_n\sqrt{2gl_1(t)}}{S} \quad (66)$$

$$\frac{dl_2(t)}{dt}=\frac{\mu_{12}S_n\text{sgn}[l_1(t)-l_2(t)]\times\sqrt{2g|l_1(t)-l_2(t)|}}{S}-\frac{\mu_{20}S_n\sqrt{2gl_2(t)}}{S} \quad (67)$$

Similarly, leakage in tank 2 can be modeled by adding a term to Equation (65) that models the leakage from tank 2. The system model for this fault is as follows:

$$\frac{dl_1(t)}{dt}=\frac{q_1(t)}{S}-\frac{\mu_{12}S_n\text{sgn}[l_1(t)-l_2(t)]\sqrt{2g|l_1(t)-l_2(t)|}}{S} \quad (68)$$

$$\frac{dl_2(t)}{dt}=\frac{\mu_{12}S_n\text{sgn}[l_1(t)-l_2(t)]\sqrt{2g|l_1(t)-l_2(t)|}}{S}-\frac{\mu_{20}S_n\sqrt{2gl_2(t)}}{S}-\frac{\mu_{20}S_n\sqrt{2gl_2(t)}}{S} \quad (69)$$

The goal of this work is to detect whether or not the system is faulty (fault detection) and, when it is faulty, to indicate the location of the fault (fault isolation). To perform this task, it is assumed that only the measurements of the water levels (l_1 and l_2), which are influenced by leakages are available. These measurements are generated by simulating the behaviour of the system in MATLAB for 3000s for three different conditions, one healthy operating condition and two faulty, in the following situations:

- The system is not faulty. The system is non-faulty from t=0 s to t=1000 s;
- The system is faulty because of leakage in tank 1. From t=1000 s to t=2000 s, there is a leakage in tank 1;
- The system is faulty because of leakage in tank 2. From t=1000 s to t=3000s , there is a leakage in tank 2;

For the first part of this study the noise in the system is assumed to have Gaussian distribution. Therefore, a zero-mean normal distribution measurement noise with standard deviation equal to 0.02 is added to both l_1 and l_2, respectively. The three models of the system are simulated in MATLAB using an ODE solver with fixed time step to get real system behavior in healthy and faulty conditions. The time step is considered to be 1s and the system is simulated for 3000 time steps from the initial conditions $l_1 = 1m$ and $l_2 = 0$. Simulation is started with healthy system and from t=1000s first fault (leak in tank 1) is introduced to the system and then at t=2000s the type of fault is changed from fault one to fault two (leak in tank 2). The input flow is assumed to be $q_1 = 1 \times 10^{-4}\mathrm{m}^3\mathrm{s}^{-1}$ in this simulation. The two tank system is generally a very slow system with the Nyquist frequency (f) around 0.1 Hz. Therefore, the sampling rate f_s is considered to be 1 Hz by knowing the Nyquist frequency in the system and using f_s=(10~20)f to calculate the desired sampling rate. The two-tank system itself acts as a low pass filter for the process noise; however, for the measurement noise the selected sampling rate may perform as a low pass filter that filters out high-frequency components of measurement noise and causes some nonlinear artifacts in the measurement noise.

As mentioned before different kind of filters (linear Kalman filter (KF), Extended Kalman filter (EKF), Unscented Kalman filter (UKF) and Particle filter (PF)) can be used to approximate the posterior state distribution in IMM-FDI method. In addition to PF-based FDI method described in Section 4.1, KF, EKF, and UKF –based FDI have also been considered for both comparison and performance evaluation for the proposed particle-filtering-based technique. Given the analytical model of two-tank system, the four filters are designed. The two-tank analytical model is being linearized around its operating point to design Kalman filter and Extended Kalman filter but for unscented kalman filter and particle filter the nonlinear model is being utilized. Level data obtained from simulation of analytical model and input flow used in simulation are being applied to estimate the water level in the tanks by IMM estimators in MATLAB.

Levels simulated by the analytical model in MATLAB and levels estimated by IMM-PF model are plotted together in Figure 5.13 and 5.14 to demonstrate the tracking performance of the designed multiple particle filter model.

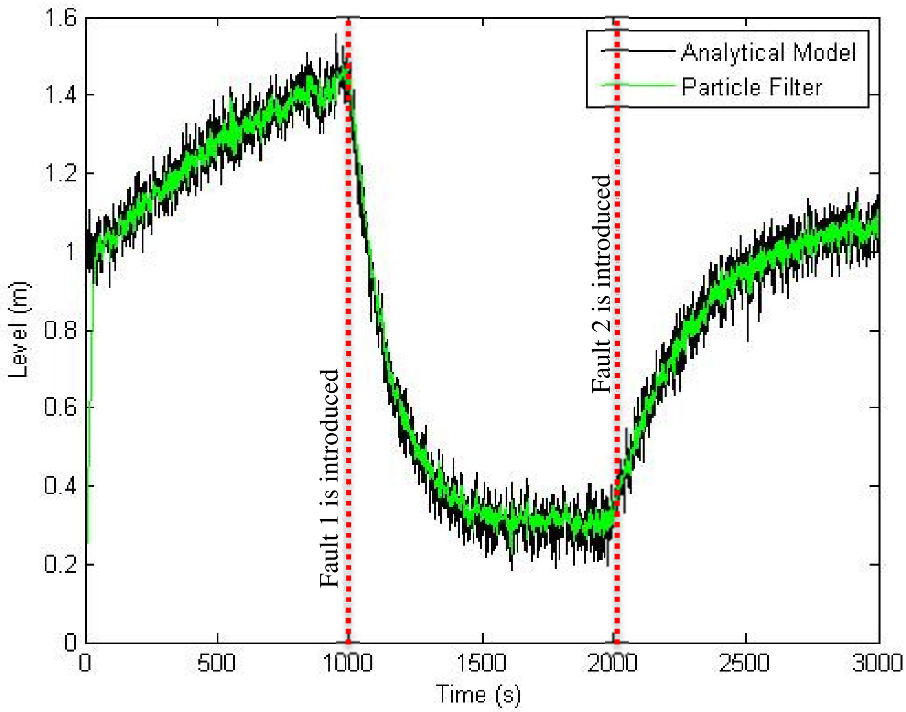

Figure 5.13: Level of tank 1 from analytical two-tank model and particle filter (Gaussian noise)

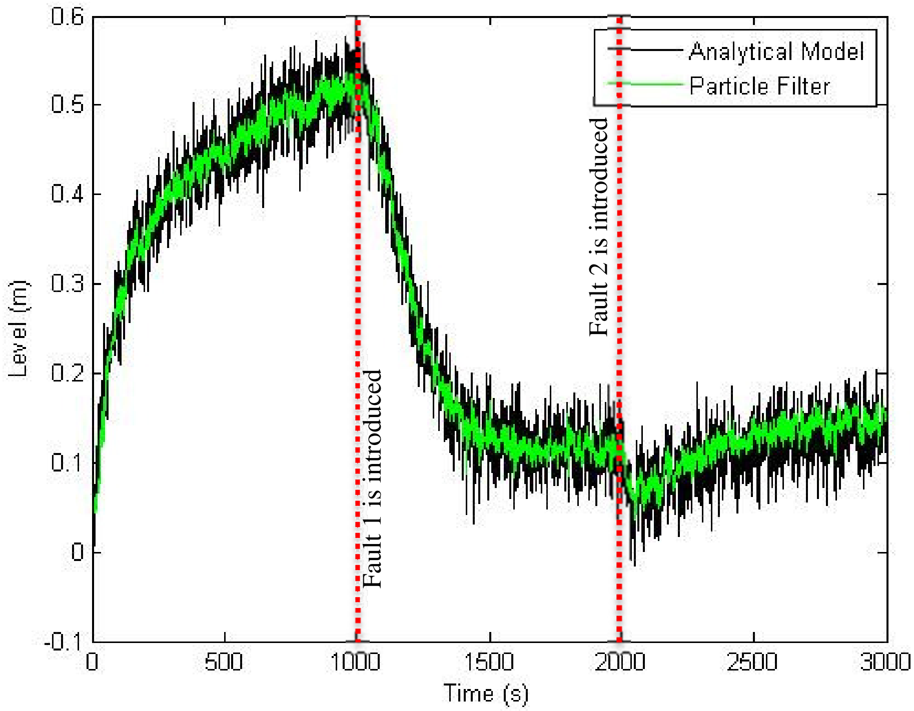

Figure 5.14: Level of tank 2 from analytical two-tank model and particle filter (Gaussian noise)

As mentioned before in section 3.1 the likelihood function can be obtained from estimated system output by IMM filter and calculating the measurement innovation. The likelihood function can be used for determining the probability of each mode of the system. Each of the applied different filter structure should be designed for three different system modes to estimate system output for each of these modes. IMM-FDI method uses innovation and covariance of innovation calculated for each mode to evaluate probability of each mode in every sampling time. After calculating different modes probability, system mode with highest probability is selected as mode which is in effect in the system at that sampling time. The results are shown in Figures 5.15 to 5.18 for different kind of filters. The red vertical lines indicate two different operating conditions of the system and the time of system changes from one mode to another mode. Mode 1, 2, and 3 represent healthy system, leak in tank 1, and leak in tank 2, respectively, in these figures.

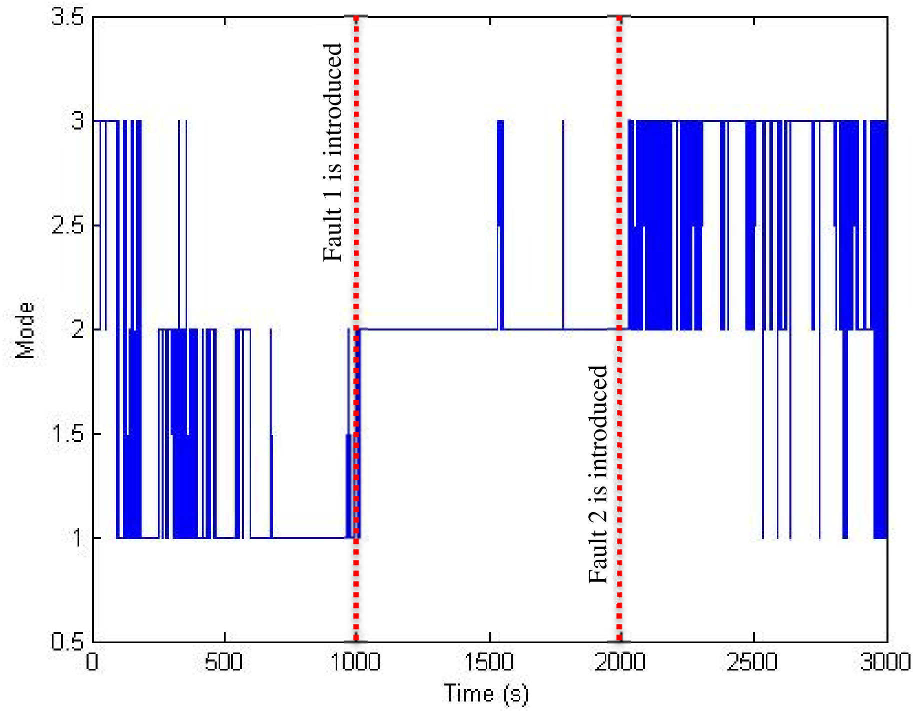

Figure 5.15: Mode is in effect in each sampling time with Kalman filter as estimator (Gaussian noise)

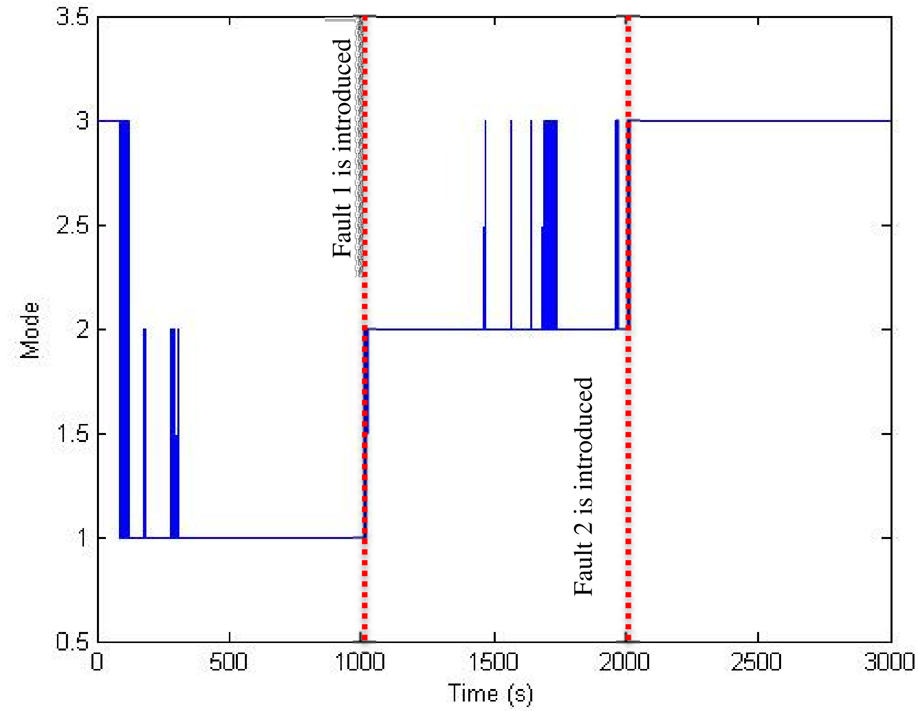

Figure 5.16: Mode is in effect in each sampling time with EKF as estimator (Gaussian noise)

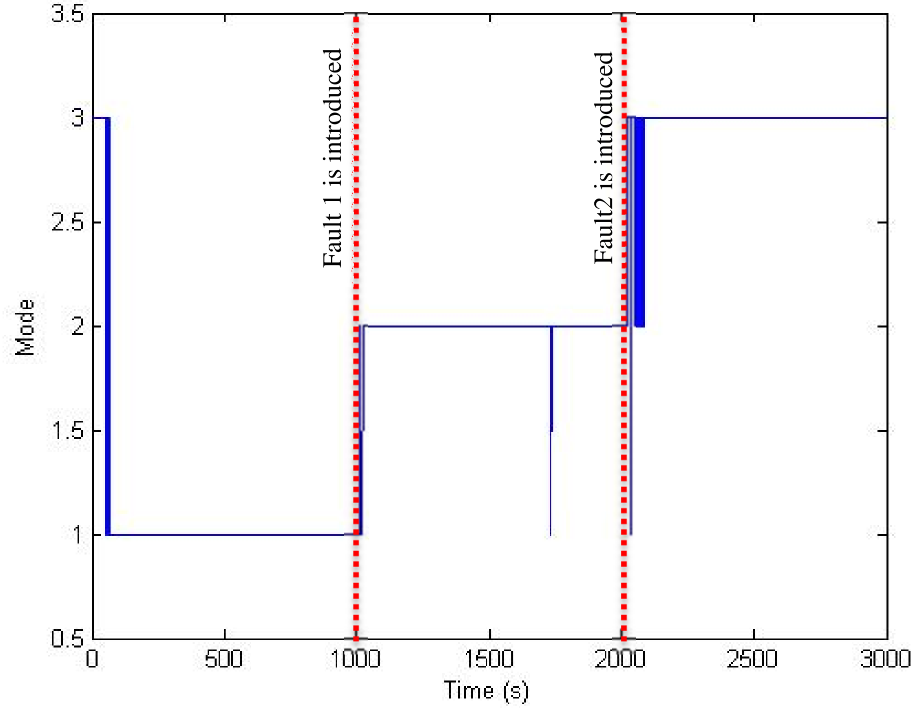

Figure 5.17: Mode is in effect in each sampling time with UKF as estimator (Gaussian noise)

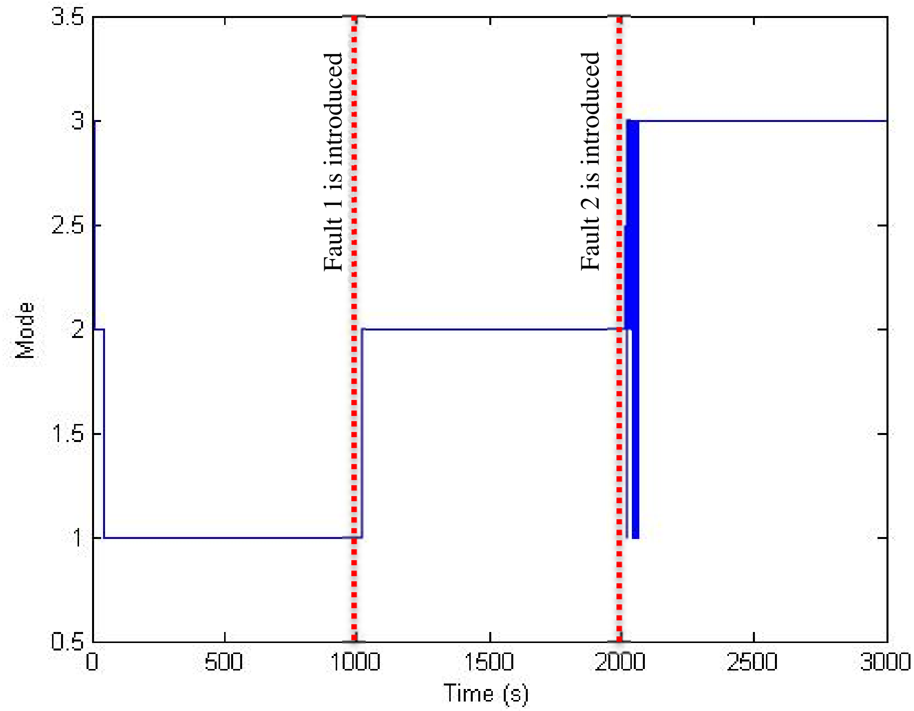

Figure 5.18: Mode is in effect in each sampling time with particle filter as estimator (Gaussian noise)

To simplify the comparison between different filters, the result of the classification for each filter is presented in Tables 5.4 to 5.7 by demonstrating confusion matrix for each of them. A simple way to define the classification accuracy of a fault diagnosis method is to calculate the confusion matrix, which summarizes the classification performance of a classifier with respect to some test data. Confusion matrix is a square array of numbers set out in rows and columns where each column of the matrix represents the instances in a predicted mode, while each row represents the instances in an actual mode. Diagonals represent correct classification according to reference data and off-diagonals represent misclassification.

Table 5.4: Confusion matrix for KF

Predicted Mode

Actual Mode	Healthy	Fault 1	Fault 2
Healthy	716	169	115
Fault 1	5	975	20
Fault 2	37	250	713

Table 5.5: Confusion matrix for EKF

Predicted Mode

Actual Mode	Healthy	Fault 1	Fault 2
Healthy	886	8	106
Fault 1	18	948	34
Fault 2	0	9	991

Table 5.6: Confusion matrix for UKF

Predicted Mode

Actual Mode	Healthy	Fault 1	Fault 2
Healthy	940	0	60
Fault 1	23	997	0
Fault 2	2	32	966

Table 5.7: Confusion matrix for PF

Predicted Mode

Actual Mode	Healthy	Fault 1	Fault 2
Healthy	958	40	2
Fault 1	17	983	0
Fault 2	12	27	961

The confusion matrix can be used for calculating different classification indices. The simplest index that can be calculated from confusion matrix is overall classification accuracy. The classification accuracy depends on the number of samples correctly classified and is calculated as following:

$$\text{classification accuracy} = \frac{\text{the number of sample cases correctly classified}}{\text{the total number of sample cases}}$$

where the number of sample cases correctly classified can be calculated by adding up all the diagonal elements in confusion matrix, and the total number of sample cases can be calculated by adding up all elements in confusion matrix. Table 5.8 shows the classification accuracy and error rate for each of the different filters.

Table 5.8: Classification accuracy (Gaussian noise)

	KF	EKF	UKF	PF
Classification accuracy	0.801	0.942	0.968	0.967

The results of simulation experiment for system with Gaussian noise show that EKF, UKF and PF demonstrate relatively good results in detecting and identifying the fault in the system. To demonstrate the advantage of particle filter-based IMM-FDI method over other type of filters (KF, EKF, UKF) for system with non-Gaussian noise, the same analysis on two-tank system is performed but with non-Gaussian noise. Noise with bimodal distribution is added to both l_1 and l_2 as measurement noise. The utilized bimodal measurement noise is a mixture of two Gaussian noise which their means are 0.05 and -0.05 and their standard deviation is 0.02 and with bimodal ratio of 1.5. In this case, the analytical model is simulated with MATLAB and IMM-FDI scheme is tested with all four filters that are used before with Gaussian noise. Similarly, the probability of each mode is evaluated and system mode with highest probability is selected. The results of each applied filter are shown in Figures 5.19 to 5.22. Moreover, the classification result for each filter is presented as confusion matrix in Tables 5.9 to 5.12.

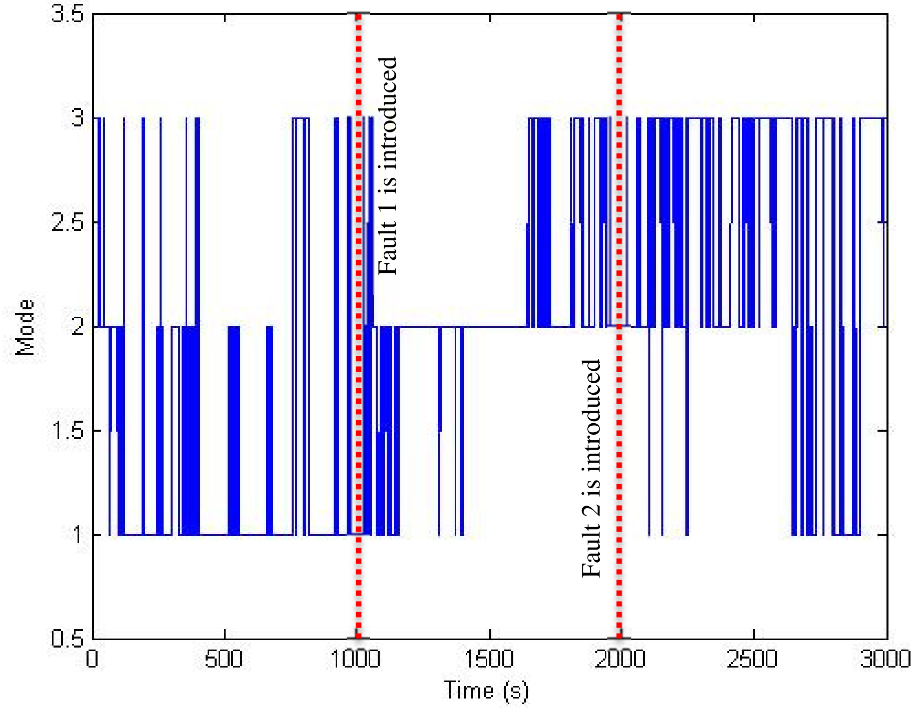

Figure 5.19: Mode is in effect in each sampling time with KF as estimator (non-Gaussian noise)

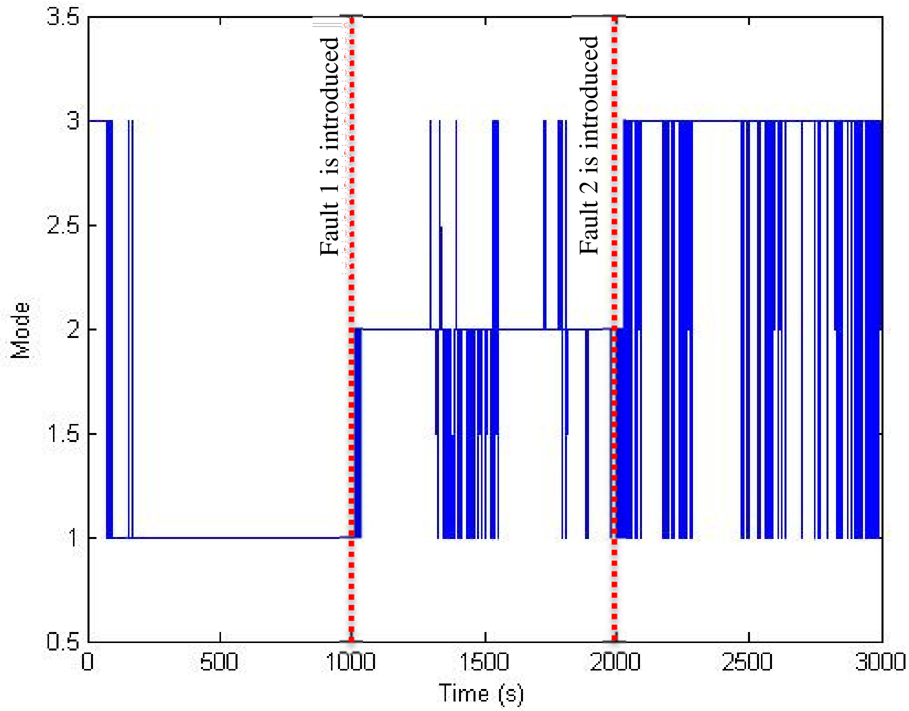

Figure 5.20: Mode is in effect in each sampling time with EKF as estimator (non-Gaussian noise)

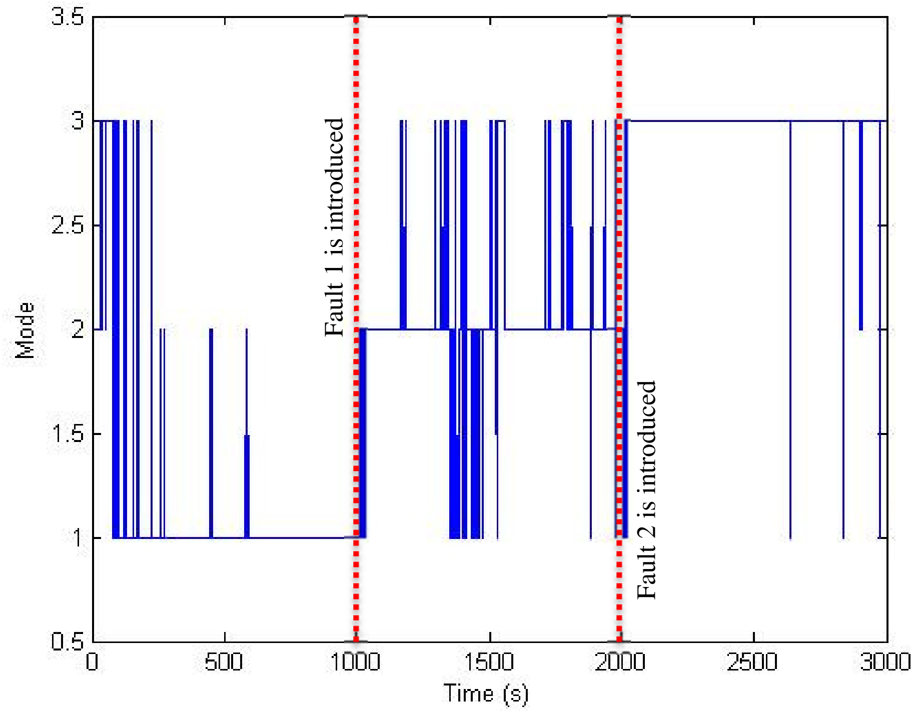

Figure 5.21: Mode is in effect in each sampling time with UKF as estimator (non-Gaussian noise)

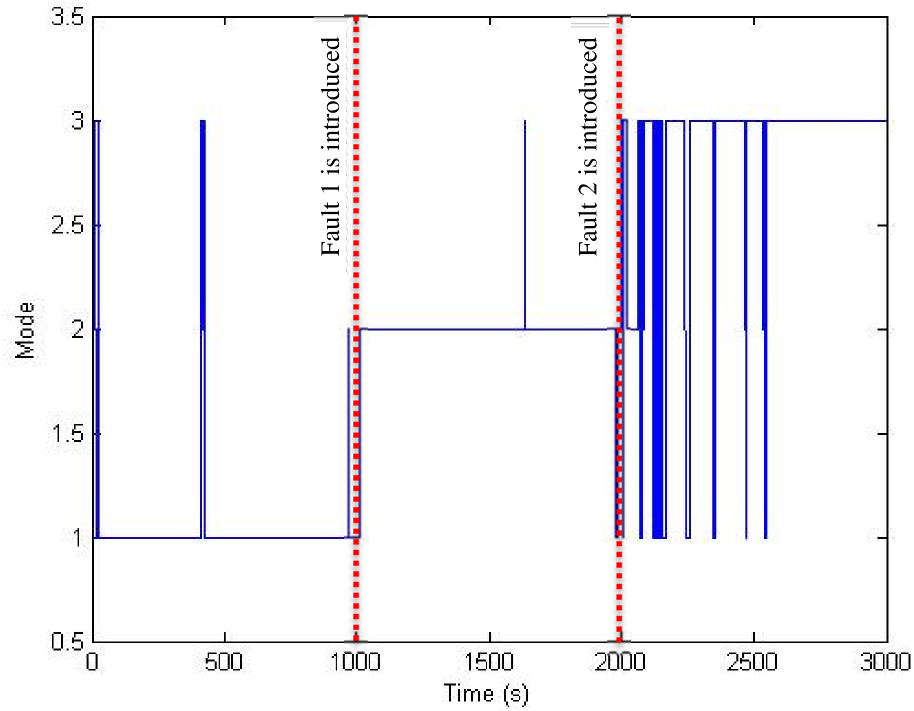

Figure 5.22: Mode is in effect in each sampling time with particle filter as estimator (non-Gaussian noise)

Table 5.9: Confusion matrix for KF

Actual Mode \ Predicted Mode	Healthy	Fault 1	Fault 2
Healthy	710	163	127
Fault 1	100	769	131
Fault 2	149	275	576

Table 5.10: Confusion matrix for EKF

Actual Mode \ Predicted Mode	Healthy	Fault 1	Fault 2
Healthy	942	0	58
Fault 1	109	976	0
Fault 2	157	10	833

Table 5.11: Confusion matrix for UKF

Actual Mode \ Predicted Mode	Healthy	Fault 1	Fault 2
Healthy	891	47	62
Fault 1	70	831	99
Fault 2	12	16	972

Table 5.12: Confusion matrix for PF

Actual Mode \ Predicted Mode	Healthy	Fault 1	Fault 2
Healthy	969	8	23
Fault 1	24	973	3
Fault 2	71	50	879

The classification accuracy is then calculated for each filter and presented in Table 5.13.

Table 5.13: Classification accuracy (non-Gaussian Noise)

	KF	EKF	UKF	PF
Classification accuracy	0.685	0.871	0.898	0.940

The results of simulation experiment and investigating the confusion matrix and accuracy rate for different filters demonstrate that particle filter performs better compared to other filters. Particle filter exhibits promising performance in detecting different faults. Also the rate of missed faults and false alarms are quite low in this method. This performance is very noticeable in terms of the classification accuracy which PF achieves higher accuracy compared to KF based techniques. This difference in classification accuracy is mainly caused by the fact that the KF-based methods perform poorly for systems with non-Gaussian noises, whereas the PF is able to handle non-Gaussian noises in the system.

5.3 Implementation on An Analytical Dynamic Suspension Model

Since the result shows that PF demonstrates better performance compared to UKF and EKF, the IMM-PF FDI method is implemented for strut suspension condition monitoring in this section.

An automotive damper is inherently non-linear due to the fluid orifice damping mechanism (Pananda et.al., 2012), although a linear viscous damping model is commonly used to represent an automotive fluid damper in the literature since this is an adequate model for their purpose (e.g. frequency analysis). However, as stated before for IMM-based FDI an accurate model of the system is very important and better model can lead to better performance in fault detection and diagnosis. Therefore, a nonlinear analytical half-car dynamic suspension model is used to simulate the strut suspension system and to test IMM-based FDI method. In this study the nonlinearity in suspension damper is modelled with a nonlinear square damping term. The half car dynamic system is elaborated in detail in section 5.1 and its schematic is shown in Figure 5.2. The road roughness is the input to the model and wheels displacement is the output of the system.

Equations 48 to 51 represent the equations of motion for the car body and the front and rear wheels. The constraints are given by

$$z_{b1} = z_b - L_1\omega \tag{70}$$

$$z_{b2} = z_b - L_2\omega \tag{71}$$

After putting z_{b1}, z_{b2} from equations (70) and (71) into equation (48),(49), (50),(51) the half car model equations are being simplified as follow:

$$m\ddot{z}_b + b_{b1}(\dot{z}_b - L_1\dot{\omega} - \dot{z}_{w1})|\dot{z}_b - L_1\dot{\omega} - \dot{z}_{w1}| + b_{b2}(\dot{z}_b - L_2\dot{\omega} - \dot{z}_{w2})|\dot{z}_b - L_2\dot{\omega} - \dot{z}_{w2}| + k_{b1}(z_b - L_1\omega - z_{w1}) + k_{b2}(z_b - L_2\omega - z_{w2}) = 0 \tag{72}$$

$$I\ddot{\omega} + L_1\big(b_{b1}(\dot{z}_b - L_1\dot{\omega} - \dot{z}_{w1})|\dot{z}_b - L_1\dot{\omega} - \dot{z}_{w1}| + k_{b1}(z_b - L_1\omega - z_{w1})\big) - L_2\big(b_{b2}(\dot{z}_b - L_2\dot{\omega} - \dot{z}_{w2})|\dot{z}_b - L_2\dot{\omega} - \dot{z}_{w2}| + k_{b2}(z_b - L_2\omega - z_{w2})\big) = 0 \tag{73}$$

$$m_{w1}\ddot{z}_{w1} - b_{b1}(\dot{z}_b - L_1\dot{\omega} - \dot{z}_{w1})|\dot{z}_b - L_1\dot{\omega} - \dot{z}_{w1}| - k_{b1}(z_b - L_1\omega - z_{w1}) + k_{w1}(z_{w1} - z_{r1}) = 0 \tag{74}$$

$$m_{w2}\ddot{z}_{w2} - b_{b2}(\dot{z}_b - L_2\dot{\omega} - \dot{z}_{w2})|\dot{z}_b - L_2\dot{\omega} - \dot{z}_{w2}| - k_{b2}(z_b - L_2\omega - z_{w2}) + k_{w2}(z_{w2} - z_{r2}) = 0 \tag{75}$$

To build the IMM model, the half car model (equations (72) to (75)) needs to be converted into the first order differential equation form. For doing this the following variables substitution is considered. These can be considered as states of the system.

$y_1 = z_b$

$y_2 = \dot{z}_b$

$y_3 = \omega$

$y_4 = \dot{\omega}$

$y_5 = z_{w1}$

$y_6 = \dot{z}_{w1}$

$y_7 = z_{w2}$

$y_8 = \dot{z}_{w2}$

And finally the suspension system equations become:

$$\dot{y}_1 = y_2 \tag{76}$$

$$\dot{y}_2 = \tag{77}$$

$$[k_{b1}(y_1 + L_1 y_3 - y_5) + k_{b2}(y_1 - L_2 y_3 - y_7) + b_{b1}(y_2 + L_1 y_4 - y_6)|y_2 + L_1 y_4 - y_6| + b_{b2}(y_2 - L_2 y_4 - y_8)|y_2 - L_2 y_4 - y_8|]/-m = 0$$

$$\dot{y}_3 = y_4 \tag{78}$$

$$\dot{y}_4 \tag{79}$$

$$= [L_1 k_{b1}(y_1 + L_1 y_3 - y_5) - L_2 k_{b2}(y_1 - L_2 y_3 - y_7) + L_1 b_{b1}(y_2 + L_1 y_4 - y_6)|y_2 + L_1 y_4 - y_6| - L_2 b_{b2}(y_2 - L_2 y_4 - y_8)|y_2 - L_2 y_4 - y_8|]/-I$$

$$\dot{y}_5 = y_6 \tag{80}$$

$$\dot{y}_6 = [k_{b1}(y_1 + L_1 y_3 - y_5) + k_{w1}(-y_5 + z_{r1}) + b_{b1}(y_2 + L_1 y_4 - y_6)|y_2 + L_1 y_4 - y_6|]/m_{w1} \tag{81}$$

$$\dot{y}_7 = y_8 \tag{82}$$

$$\dot{y}_8 = [k_{b2}(y_1 - L_2 y_3 - y_7) + k_{w2}(-y_7 + z_{r2}) + b_{b2}(y_2 - L_2 y_4 - y_8)|y_2 - L_2 y_4 - y_8|]/m_{w2} \tag{83}$$

It is assumed that y_5 and y_7, displacement of front and rear wheels, respectively, are measureable in the system. And the output equation is as follow:

$$[z_1 \quad z_2] = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} \tag{84}$$

To generate a component fault in the course of a simulation, one way is to build a new model, whose components are already faulty. Therefore, to apply IMM-FDI scheme, the half car model suspension system should be modeled separately for each of its working conditions *e.i.* healthy and faulty working conditions. These models are used to generate the wheel's displacement signals and different faults in strut.

Two faults are considered in this study.

1. Internal leakage through the piston seal in strut which causes increase in the damping effect of the strut and it is modelled by bigger damping coefficient
2. Broken control valve in strut which causes decrease in damping effect of the strut and it is modelled by smaller damping coefficient

Thus, three main operating conditions are distinguished: the *normal* condition reflects the fact that the strut is healthy, meanwhile the two *faulty* conditions that indicate two faults explained above. Therefore simulation is taken in three situations of the damper condition with three different damping coefficients to model healthy system and two above mentioned faults. That is to say, three different situations are studied when the vehicle runs on with healthy strut, strut with internal leakage and strut with broken control valve. To generate measurement data associated with displacement of front and rear wheels, the half-car dynamic suspension model is simulated in MATLAB subjected to road excitation with nominal parameter values presented in table 5.1 (Gao et al., 2007). Simulation test lasts for 72 seconds during which internal leakage fault is introduced into the healthy system at the 24th second and broken control valve fault is introduced to the system at 48th second. The vehicle velocity is assumed to be 10 km/h. Three bumps of height 2cm are considered on the road in this simulation which each of these bumps are exciting the suspension system in different operating condition.

It is also assumed that only wheel's displacement is measurable and the measurement is contaminated by a noise with bimodal distribution. The utilized bimodal measurement noise is a mixture of two Gaussian noises which their means are 0. 0003 and -0. 0003 and their standard deviation is 0. 0001 and with bimodal ratio of 1.5. Figures 5.23 and 5.24 show the displacement of the front and rear wheels, respectively. As shown in Figures 5.23 and 5.24, it is observable that fault one causes smaller excitation because of bump in the car body and it is dampen away faster whoever fault two results bigger displacement due to bump with more oscillation than normal.

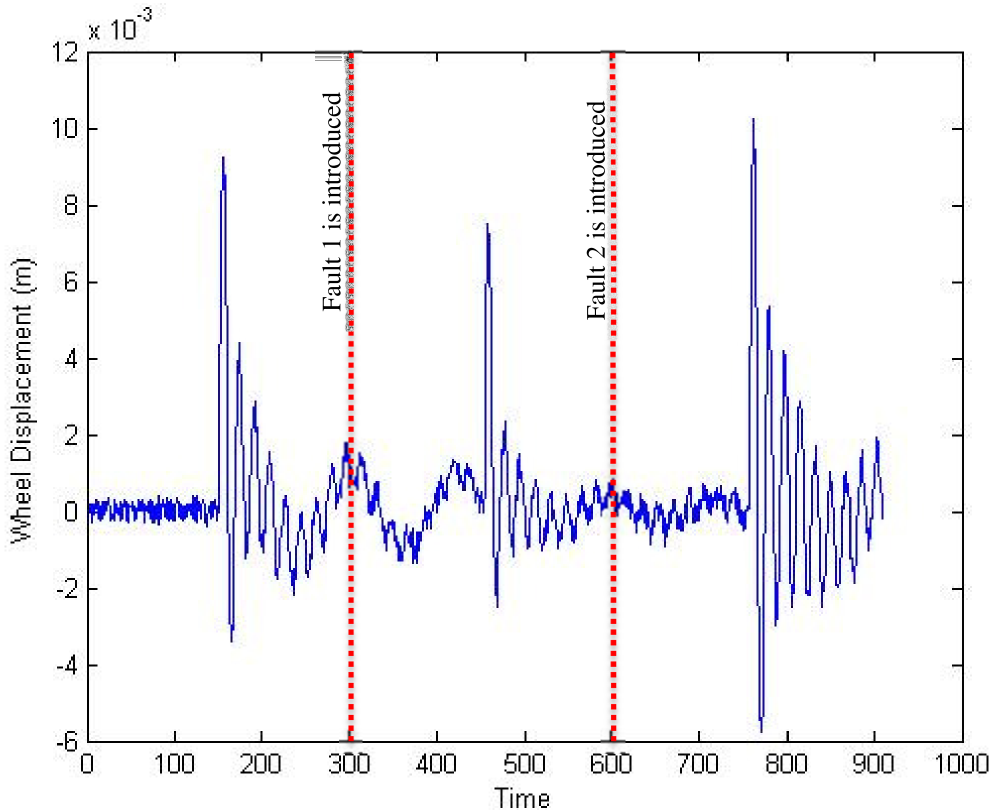

Figure 5.23: Front wheel displacement

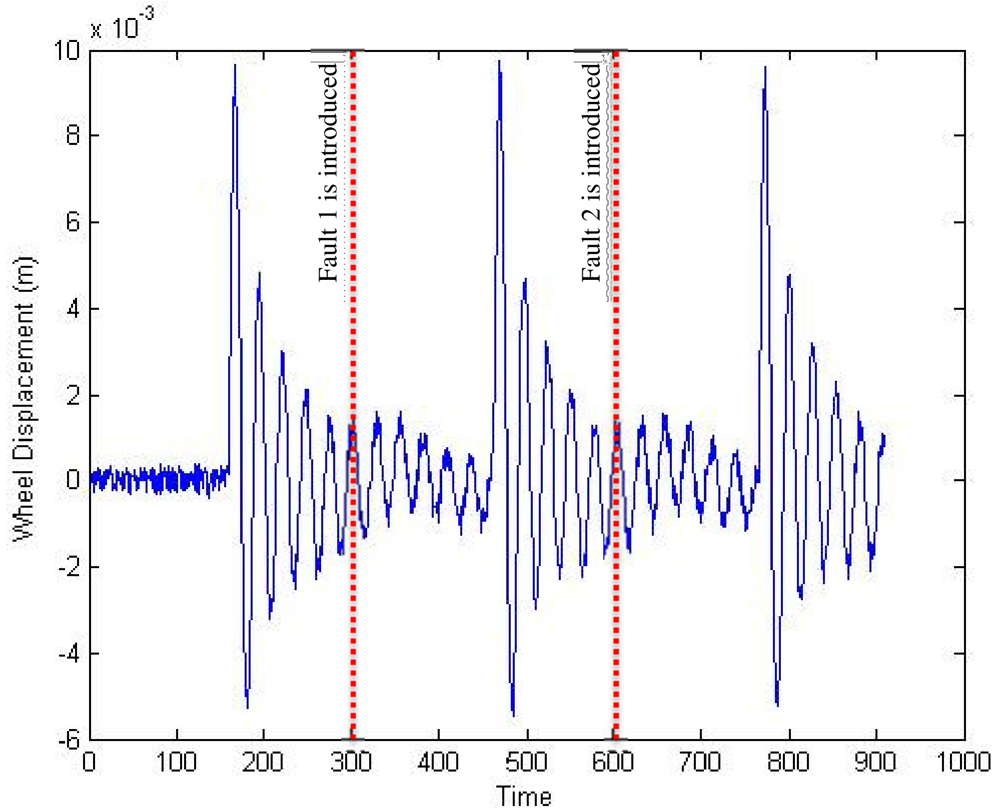

Figure 5.24: Rear wheel displacement

Given that noise in the system is assumed to have bimodal distribution, particle filter is applied as estimator in IMM approach for this study. Figure 5.25 shows the results obtained when the proposed FDI approach is applied to simulation data from strut model in different operating conditions. It indicates the mode of the system with biggest probability at each sampling time. The red vertical lines discriminate between two different operating conditions of the system and they show the time that system switches from one mode to another mode. In these figures mode 1, 2, and 3 represent healthy system, internal leakage, and broken control valve, respectively.

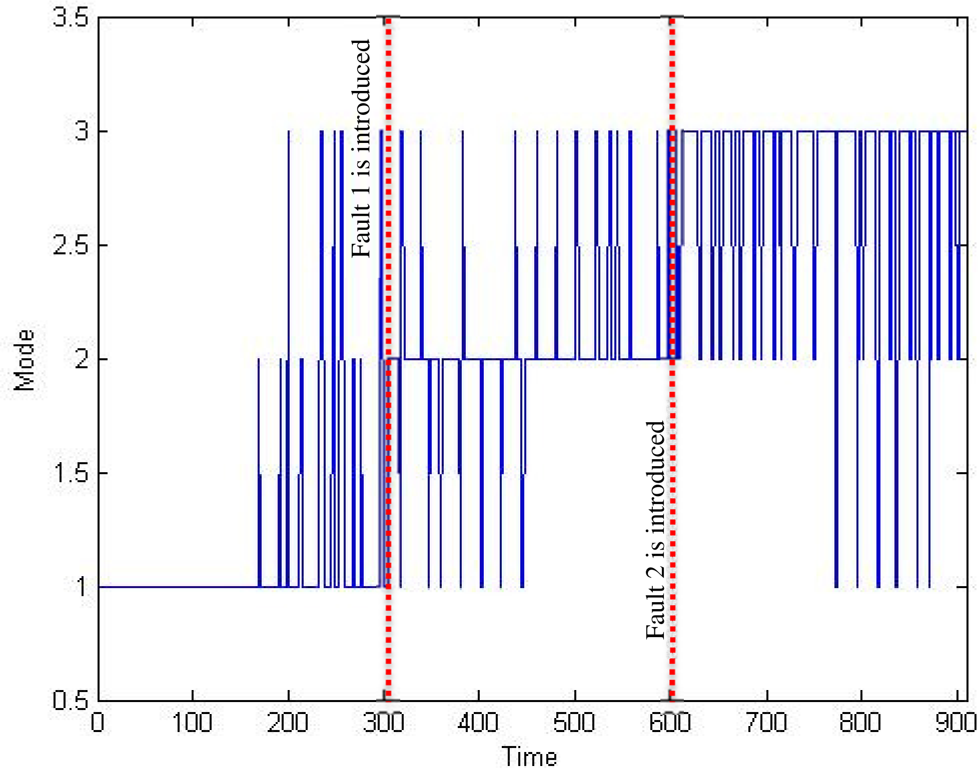

Figure 5.25: Mode is in effect in each sampling time

Table 5.14: Confusion matrix

Actual Mode \ Predicted Mode	Healthy	Fault 1	Fault 2
Healthy	278	11	14
Fault 1	13	265	25
Fault 2	12	49	242

The performance of this FDI method as confusion matrix is shown in Table 5.14 and the classification accuracy is calculated to be 0.863.

As it is obvious in Figure 5.25, there are some misclassification during the classification process of the system. To reduce the misclassification rate of the FDI method, the mode calculated by IMM filter is feed through a 4 points moving average filter. The results of classification are shown in Figure 5.26.

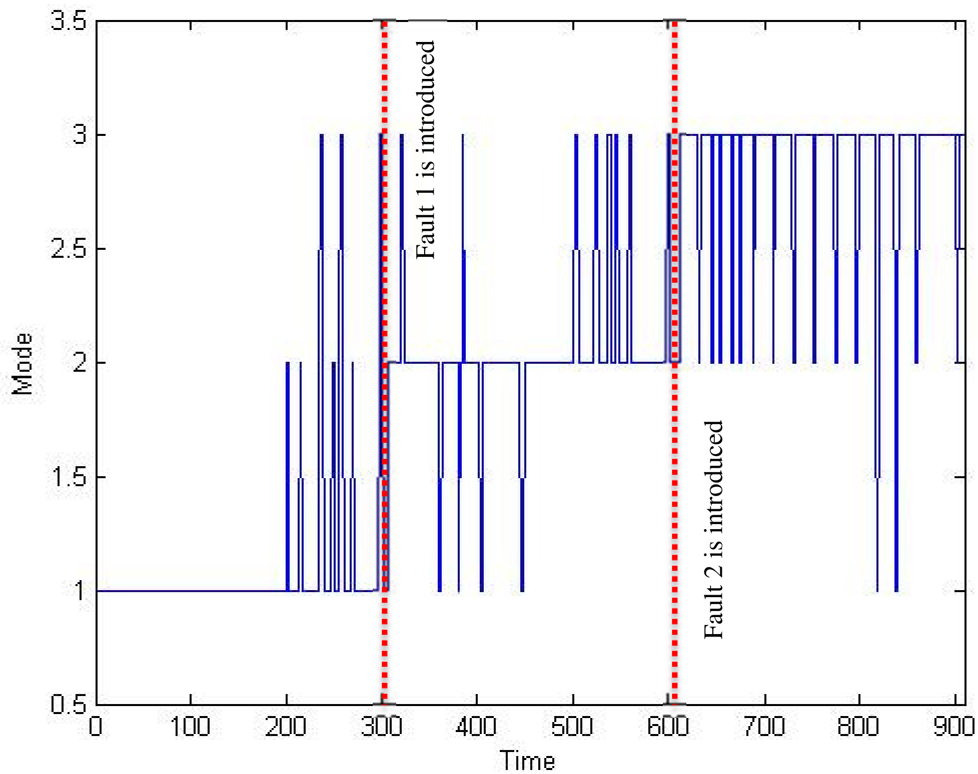

Figure 5.26: Mode is in effect in each sampling time after applying moving average filter

Also, the classification performance is illustrated in Table 5.15. The classification accuracy after applying the moving average filter is increased to 0.883.

Table 5.15: Confusion matrix after applying moving average filter

Actual Mode \ Predicted Mode	Healthy	Fault 1	Fault 2
Healthy	279	16	8
Fault 1	14	265	24
Fault 2	4	40	259

As shown in Figure 5.26, the moving average filter improves the classification accuracy but still there are some misclassification observable in this figure. One way to avoid these misclassification is to define appropriate threshold and if the probability for a specific faulty mode exceeds the threshold for a number of times the alarm indicator would be activated.

To study the effect of noise strength in performance of FDI method the model is simulated with the same type of noise but different intensity. In this part the means are 0.0006 and -0.0006 and the standard deviation is 0.00025 and with bimodal ratio of 1.5. The same structure of IMM model with the same number of particles in particle filter is implemented. The results are shown in Table 5.16 and classification accuracy for this analysis is 0.625. It can be concluded that the IMM-PF method shows poor classification results for system with bigger noise intensity. One way to overcome this problem is to increase the number of particles which causes longer computational time. So there is always a compensation between the performance of FDI method and available computational power.

Table 5.16: Confusion matrix for system with bigger noise intensity

Predicted Mode

Actual Mode		Healthy	Fault 1	Fault 2
	Healthy	217	40	46
	Fault 1	44	191	68
	Fault 2	28	115	160

Chapter 6: Conclusions and Future Work

6.1 Conclusions

In this research fault detection method for nonlinear systems is studied. Condition monitoring of mining haul truck suspension strut was the focus of study in this research. Initially, data-driven fault detection method is applied for FDI of strut by utilizing strut pressure as observable variable. Combination of continuous wavelet transform and PCA is used to handle the task of fault detection for strut. By application of the wavelet transform, it was possible to extract some features in fault categories to determine specific boundaries for each group of faults, and to classify the probable faults in the system according to the corresponding boundary. PCA is used for ranking the feature extracted by CWT. The nonlinear nature of the process and also non-Gaussian noises in the system leads to poor results in applied data-driven FDI approach. Also, the applied data-driven method was not able to do fault identification and fault diagnosing. Therefore, a hybrid model-based FDI method is applied on suspension strut to have more reliable FDI method . An interacting multiple model fault detection approach for fault detection and diagnosis is applied on hydraulic two-tank system as benchmark problem with different filters, *e.g.* KF, EKF, UKF, PF. These different filters are used to construct IMM structure and to compare their performance with nonlinear stochastic system in IMM structure. The results show that PF performs better in fault detection and diagnosis of nonlinear stochastic systems and it can tackle nonlinearities and non-Gaussian noises in the system. Lastly, the IMM-based FDI method is applied on an analytical half car model to investigate its performance on the suspension strut. Two different faults are considered in this study for half car model and it is being simulated for three different situations (healthy and two faulty situation). The results are reported as confusion matrix and classification accuracy. These results show that IMM-PF demonstrate promising results in fault detection and diagnosis of suspension strut. The results can be improved by applying a predefined threshold for probability of occurrence of each specific fault in the system and compare the calculated probability for each fault with corresponding threshold. This can help to reduce the rate of misclassification and false alarms.

6.2 Future Work

For further study the half car model system and corresponding parameters can change in a way to simulate the real mining truck as close as possible. Another limitation in the modeling process is the road pattern that might not completely simulate the road profile in the mine. In the future study a displacement sensor can be installed on one haul truck in the fleet to log the road profile to provide better understanding of the irregularities and bumps in the mine haul roads. Road roughness (i.e. input to the system) is not measurable in the actual strut problem. Systems that some or all of the inputs to the system are completely unknown are called systems with unknown inputs. Regular estimators or filters cannot be used for state estimation of these systems, and special structure of them called unknown input observer should be used (Darouach, et al. 1994). One subject to study in the future would be adapting an unknown input observer to IMM-FDI structure. Another future work includes application of other kind of particle filters such as Rao-Blackwellized particle filter to improve the performance of IMM FDI method. Finally, another subject to study in the future would be application of the proposed method to a nonlinear system with variable transition matrix.

References

Alrowaie, F., Gopaluni, R. B., Kwok, K. E. (2012). Fault detection and isolation in stochastic non-linear state-space models using particle filters. Control Engineering Practice, 20(10), 1016-1032.

Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, Signal Processing. IEEE Transactions on, 50 (2), 174-188.

Banerjee, A., Arkun, Y., Ogunnaike, B., Pearson, R. (1997). Estimation of nonlinear systems using linear multiple models. AIChE Journal, 43(7), 1204-1226.

Bar-Shalom, Y., Fortmann, T. (1988). Tracking and data association. Mathematics in Science and Engineering, 179. Academic Press.

Beard, R. V. (1971). Failure accommodation in linear systems through self-organization. Man Vehicle Laboratory, MIT, Cambridge.

Bhagwat, A., Srinivasan, R., Krishnaswamy, P. R. (2003a). Fault detection during process transitions: A model-based approach. Chemical Engineering Science, 58(2), 309-325.

Bhagwat, A., Srinivasan, R., Krishnaswamy, P. R. (2003b). Multi-linear model-based fault detection during process transitions. Chemical Engineering Science, 58(9), 1649-1670.

Bongers, D. R., Gurgenci, H. (2008). Fault detection and identification for Longwall machinery using SCADA data. Complex system maintenance, Springer Series in Reliability Engineering, Part F, 611-641.

Börner, M., Straky, H., Weispfenning, T., Isermann, R. (2000). Model based fault detection of vehicle suspension and hydraulic brake systems.1st IFAC Conference on Mechatronic systems, Darmstadt, Germany.

Bruno, M. G. S. (2013). Sequential monte carlo methods for nonlinear discrete-time filtering. Synthesis lectures on signal processing, 11(1), 1-99.

Chiang, L., Russell, E., Braatz, R. (2001). Fault detection and diagnosis in industrial systems. Springer, Nueva York.

Chen, J., Patton, R. J. (1999). Robust model-based fault diagnosis for dynamic systems. Kluwer Academic Publishers, Norwell, Mass.

Chen, M. Z., Zhou, D. H., Liu, G. P. (2005). A new particle predictor for fault prediction of nonlinear time-varying systems. Developments in Chemical Engineering and Mineral Processing, 13(3-4), 379-388.

Chow, E., Willsky, A. (1984). Analytical redundancy and the design of robust failure detection systems. IEEE Trans, on Automatic Control, 29(7), 603- 614.

Cui, N., Hong, L., Layne, J. R. (2005). A comparison of nonlinear filtering approaches with an application to ground target tracking, Signal Processing, 85 (8), 1469-1492.

Dalle Molle, D. T., Himmelblau, D. M. (1987). Fault detection in a single-stage evaporator via parameter estimation using the Kalman filter. Industrial & Engineering Chemistry Research, 26, 2482–2489.

Darouach, M., Zasadzinski, M., Xu, S. J. (1994). Full-order observers for linear systems with unknown inputs. IEEE Transactions on Automatic Control, 39(3), 606-609.

Ding, S. X. (2007). Model based fault diagnosis technique. Springer Verlag, Berlin.

Dixon, J. C. (2007). The shock absorber handbook. John Wiley & Sons, Ltd, Chichester, UK.

Dodds, C. J., Robson, J. D. (1973). The description of road rurface roughness. J. of Sound and Vibration, 31(2), 175-183.

Donders, S. (2004). Fault detection and identification for wind turbine systems: A closed-loop analysis. Master's Thesis, University of Twente.

Dong, D., McAvoy, T. J. (1996). Batch tracking via nonlinear principal component analysis. AIChE J., 42(8), 2199-2208.

Doucet, A., De Freitas, N., Gordon, N. (2001). Sequential Monte Carlo Methods in Practice. Springer-Verlag, New-York.

Dunia, R., Qin, S. J., Edgar, T. F., McAvoy, T. J. (1996). Identification of faulty sensors using principal component analysis. AIChE J., 42(10), 2797-2812.

Efe, M., Atherton, D. P. (1997). The IMM approach to the fault detection problem. Conference Proceedings SYSID '97, 2, 625–630.

Fathi, Z., Ramirez, W. F., Korbicz, J. (1993). Analytical and knowledge-based redundancy for fault diagnosis in process plants. AIChE J., 39(1), 42–56.

Fekih, A., Xu, H., Chowdhury, F. N. (2007). Neural networks based system identification techniques for model based fault detection of nonlinear systems. International Journal of Innovative Computing, Information and Control, 3(5), 1073-1085.

Fisher, D., Kaus, E., Isermann, R. (2003). Fault detection for an active vehicle suspension. Proceeding of the American Control Conference, Denver, Colorado.

Frank, P. M. (1987). Fault diagnosis in dynamic systems via state estimation – A survey. In Tzafestas, S. G. et al. (Eds), System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches, 1, 35-98.

Freeman, P., Pandita, R., Srivastava, N., Balas, G. J. (2013). Model-based and data-driven fault detection performance for a small UAV. IEEE/ASME Transactions on Mechatronics.

Gadsden, S. A., Habibi, S. R. (2011). Model-based fault detection of a battery system in a hybrid electric vehicle. Paper presented at the 2011 IEEE Vehicle Power and Propulsion Conference, VPPC 2011.

Galan, O., Romagnoli, J. A., Palazoglu, A. (2000). Robust H infinity control of nonlinear plants based on multi-linear models: An application to a bench-scale pH neutralization reactor. Chemical Engineering Science, 55, 4435–4450.

Gao, W., Zhang, N., Du, H. P. (2007). A half-car model for dynamic analysis of vehicles with random parameters. 5th Australasian Congress on Applied Mechanics, ACAM, Brisbane, Australia.

García-Álvarez, D. (2009). Fault detection using Principal Component Analysis (PCA) in a Wastewater Treatment Plant (WWTP). 62th International Students Scientific Conference.

Gelb, A. (1974). Applied optimal estimation. Cambridge, MA: MIT Press.

Gertler, J. (1998). Fault detection and diagnosis in engineering systems. Marcel Dekker Inc.

Gertler, J. J., Costin, M., Fang, X., Hira, R., Kowalczuk, Z., Luo, Q. (1993). Model-based on-board fault detection and diagnosis for automotive engines. Control Engineering Practice, 1(1), 3-17.

Goda, K., Goodall, R. M. (2004). Fault detection and isolation system to a railway vehicle bogie, vehicle system dynamics. Supplement 41, 468-476.

Goffaux, G., Wouwer, A. (2005). Bioprocess state estimation: Some classical and less classical approaches. Control and observer design for nonlinear finite and infinite dimensional systems, Lecture Notes in Control and Information Sciences, 322, 111-128.

Goodall, R. (2006). Advanced control and monitoring for railway suspensions. KRRI Seminar, London.

Hajizadeh, M., Lipsett, M. G. (2013). Anomaly detection in mining haul truck suspension struts. International Journal of Condition Monitoring (accepted with minor revision and revised revision is submitted).

Handschin, J. E., Mayne, D. Q. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. International Journal of Control, 9, 547–559.

Harvey L. H. (1977). A study of the characteristics of automotive hydraulic dampers at high stroking frequencies. Ph.D. Thesis, Department of Mechanical Engineering, University of Michigan, USA.

Hayashi, Y., Tsunashima, H., Marumo, Y. (2006). Fault detection of railway vehicles using multiple model approach. SICE-ICASE, International Joint Conference, 2812-2817.

Hengy, D., Frank, P.M. (1986). Component failure detection via nonlinear state observers. IFAC Workshop, Kyoto (Japan).

Hongxing, W., Zengpu, X., Congling, Z., Lin, Y. (2008). Method for extraction wavelet packets' coefficients in loudspeaker fault detection based on PCA. Paper presented at the Proceedings-2008 Pacific-Asia Workshop on Computational Intelligence and Industrial Application, PACIIA, 1, 864-868.

Himmelspaeh, T. (1992). A model based sensor fault detection and isolation approach for smooth deterministic nonlinear dynamic systems with application. Ph.D. Thesis, University of Michigan., USA.

Hoskins, J. C., Kaliyur K. M., Himmelblau, D. M. (1991). Fault diagnosis in complex chemical plants using artificial neural networks. AIChE J., 37(1), 137—141.

Isermann, R. (1984). Process fault detection based on modeling and estimation methods. Automatica, 20(4), 387-404.

Isermann, R. (1994). Fault diagnosis via parameter estimation and knowledge processing. Automatica, 29(4), 815-835.

Isermann, R. (2006). Fault diagnosis systems: An introduction from fault detection to fault tolerance. Springer Verlag, Berlin Heidelberg.

Isermann, R., Balle, P. (1997). Trends in the application of model-based fault diagnosis of technical processes. Contr. Eng. Practice, 5, 707–719.

Isermann, R, Börner, M., Fischer, D. (2002). Mechatronic semi-active vehicle suspension. 1st International Symposium on Mechatronics, Chemnitz, Germany.

Johansen, T. A., Foss, B. A. (1999). Multiple model approaches to modeling and control. International Journal of Control, 72(7,8), 575.

Jones, H. L. (1973). Failure detection in linear systems. MIT, Cambridge..

Kadirkamanathan, V., Li, P., Kirubarajan, T. (2001). Sequential monte carlo filtering vs. the IMM estimator for fault detection and isolation in nonlinear systems. Paper presented at the Proceedings of SPIE-the International Society for Optical Engineering, 4389. 263-274.

Kadirkamanathan, V., Li, P., Jaward, M. H., Fabri, S. G. (2002). Particle filtering-based fault detection in non-linear stochastic systems. International Journal of Systems Science, 33(4), 259-265.

Khan, A. Q., Ding, S. X., Chihaia, C. I., Abid, M., Chen, W. (2010). Robust fault detection in nonlinear systems: A three-tank benchmark application. Conference on Control and Fault Tolerant Systems Nice, France.

Kim, S., Choi, J., Kim, Y., (2008). Fault detection and diagnosis of aircraft actuators using fuzzy-tuning IMM filter. IEEE Transactions on Aerospace and Electronic Systems, 44(3), 940-952.

Kinnaert, M., Vrancic, D., Denolin, E., Juricic, D., Petrovcic, J. (2000). Model-based fault detection and isolation for a gas-liquid separation unit. Control Engineering Practice, 8(11), 1273-1283.

Krishnaswami, V., Luh, G.C., Rizzoni, G. (1995). Non-linear parity equation based residual generation for diagnosis of automotive engine faults. Contr. Eng. Practice., 3(10), 1385–1392.

Li, P, Goodall, R. M. (2004). Model-based condition monitoring for railway vehicle system. Procedings of the UKACC International Conference on Control, University of Bath, UK.

Li, P., Goodall, R., Weston, P., Ling, C.S., Goodman, C., Roberts, C. (2007). Estimation of railway vehicle suspension parameters for condition monitoring. Control Engineering Practice, 15(1), 43-55.

Li, P., Kadirkamanathan, V. (2001). Particle filtering based likelihood ratio approach to fault diagnosis in nonlinear stochastic systems. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 31(3), 337-343.

Li, P., Kadirkamanathan, V. (2004). Fault detection and isolation in non-linear stochastic systems - A combined adaptive monte carlo filtering and likelihood ratio approach. International Journal of Control, 77(12), 1101-1114.

Li, R. (1991). Fault detection and diagnosis: Extended Kalman filter and neural networks. Ph.D. Thesis, Department of Chemical Engineering, University of Delaware, USA.

Li, R., Olson, J. H. (1991). Fault detection and diagnosis in a closed loop non-linear distillation process: Application of extended Kalman flters. Industrial & Engineering Chemistry Research, 30, 898–908.

Li, T., Guo, L., Wu, L. (2008). Observer-based optimal fault detection using PDFs for time-delay stochastic systems. Nonlinear Anal., 9(5), 2337–2349.

Li, W., Yue, H. H., Valle-Cervantes, S., Qin, S. J.(2000). Recursive PCA for adaptive process monitoring. J. of Process Contr., 10, 471–486.

Li, X. R., Bar-Shalom, Y. (1993). Design of an interacting multiple model algorithm for air traffic control tracking. IEEE Trans. Contr. Syst. Technol., 1, 186-194.

Lin, J., Qu, L. (2000). Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. Journal of Sound and Vibration, 234(1), 135-148.

Lin, J., Zuo, M.J. (2003). Gearbox Fault Diagnosis using Adaptive Wavelet Filter. Mechanical Systems and Signal Processing, 17(6), 1259–1269.

Luo, R., Misra, M. , Qin, S. J., Barton, R., Himmelblau, D. M. (1999). Sensor fault detection via multiscale analysis and dynamic PCA. Ind. Eng. Chem. Res., 37, 1489-1495.

MacGregor, J. F., Kourti, T. (1995) Statistical process control of multivariate processes. Contr. Eng. Practice 3(3), 403–414.

Magill, D. T. (1965). Optimal adaptive estimation of samples stochastic processes. IEEE Trans. on Automat. Contr., (4), 434-439.

Majjad, R. (1997). Estimation of suspension parameters. IEEE International Conference on Control Applications, Hartford, CT.

Mehra, R. K., Peschon, J. (1971). An innovations approach to fault detection and diagnosis in dynamic systems. Automatica, 7(5), 637–640.

Mehra, R., Rago, C., Seereeram, S. (1998). Autonomous failure detection, identification and fault-tolerant estimation with aerospace applications. Paper presented at the IEEE Aerospace Applications Conference Proceedings, 2, 133-138.

Mei, T. X., Ding, X. J. (2007). A model-less technique for the fault detection of rail vehicle suspension. 20th IAVSD Symposium, Berkeley, California, USA.

Moncada, H. B., Marin, J. A. (2011). Fault detection for a bus suspension model using an Utkin observer. Electronics, Robotics and Automotive Mechanics Conference (CERMA), IEEE , 246-251.

Morello, L., Rossini, L. R., Pia, G., Tonoli, A. (2011). Mechanical engineering series. The Automotive Body, Volume II: System Design. Springer, Netherlands. 239-363.

Ning, L., Zhang, Y., Wen, D. (2010). Industrial process fault detection based on wavelet denoising PCA. Paper presented at the 4th International Symposium on Computational Intelligence and Industrial Applications, ISCIIA, 134-138.

Nomikos, P., MacGregor, J. (1994). Monitoring batch processes using multiway principal component analysis. AIChE J., 40(8), 1361-1375.

Orchard, M. E., Vachtsevanos, G. J. (2009). A particle-filtering approach for on-line fault diagnosis and failure prognosis. Transactions of the Institute of Measurement and Control, 31(3-4), 221-246.

Özgönenel, O., Önbilgin, G., Kocaman, C. (2005). Transformer protection using the wavelet transform. Turk J Elec Engin, 13(1).

Pananandа, N., Ferguson, N. S., Waters, T. P. (2012). The effect of cubic damping in an automotive vehicle suspension model. International symposium on the computational modelling and analysis of vehicle body noise and vibration, Brighton, GB.

Patton, R. J., Chen, J., Lopez-Toribio, C. J. (1998). Fuzzy observers for nonlinear dynamic systems fault diagnosis. Proc., 37th IEEE Conf. on Decision and Control, 1, 84–89.

Pertew, A. M., Marquez, H. J., Zhao, Q. (2007). LMI-based sensor fault diagnosis for nonlinear Lipschitz systems. Automatica, 43(8), 1464-1469.

Polycarpou, M. M., Helmicki, A. J. (1995). Automated fault detection and accommodation: A learning systems approach. IEEE Trans. Syst. Man Cybern., 25(11), 1447-1458.

Qin, S. J. (1998). Recursive PLS algorithms for adaptive data modeling. Computers and Chem. Eng., 22, 503-514.

Qin, S., Li, W. (1999). Detection, identification, and reconstruction of faulty sensors with maximized sensitivity. AICHE J., 45(9), 1963-1976.

Qin, S. J., McAvoy, T. J. (1992). Nonlinear PLS modeling using neural networks. Computers and Chem. Eng., 16(4), 379-391.

Raghappriya, M., Kanthalakshmi, S., Manikandan, V. (2012). Diagnosis of faults using IMM estimator. ARPN Journal of Engineering and Applied Sciences, 7(6), 780-786.

Raich, A., Cinar, A. (1996). Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. AICHE J. 42(4), 995-1009.

Rigatos, G. G. (2009). Fault detection and isolation based on fuzzy automata. Information Sciences, 179(12), 1893-1902.

Rill, G. (2012). Road vehicle dynamics: Fundamentals and modeling. CRC Press, Taylor and Francis group. London.

Ristic, B., Arulampalam, S., Gordon N. (2004). Beyond the Kalman filter: Particle filters for tracking applications. Artech House, Boston.

Ru, J., Li, X. R. (2008). Variable-structure multiple-model approach to fault detection, identification, and sstimation. IEEE transaction on control systems technology, 16(5), 1029-1038.

Samy, I., Gu, D. W. (2011). Fault detection and flight data measurement: Demonstrated on unmanned air vehicles using neural networks. Springer, New York..

Seliger, R., Frank, P. M. (1991). Robust component fault detection and isolation in nonlinear dynamic systems using nonlinear unknown input observers. Proceedings of the IFACIIMACS Symposyium on Fault Detection, Supervision and Safety for Technical Processes- SAFEPROCESS 91, Baden-Baden, Germany, 313-318.

Silva, L., Delarmelina, D., Junco, S., M'Sirdi, N.K., Noura, H. (2007). Bond graph based fault diagnosis of 4W-vehicles suspension systems I: Passive suspensions, society for computer simulation. Journal of Simulation Series, 39(1), 217-224.

Sobhani-Tehrani, E. (2008). Fault detection, isolation, and identification for nonlinear system using a hybrid approach. Ph.D. Thesis, Concordia University, Canada.

Sobhani-Tehrani, E., Khorasani, K. (2009). Fault diagnosis of nonlinear systems using a hybrid approach. Springer, New York.

Souibgui, F., BenHmida, F., Chaari, A. (2011). Particle filter approach to fault detection and isolation in nonlinear systems. Paper presented at the International Multi-Conference on Systems, Signals and Devices, SSD'11 - Summary Proceedings.

Stribrsky, A., Hyniova, K., Honcu, J., Kruczek, A. (2003). Using fuzzy logic to control active suspension system of one-half-car model. Department of Control Engineering, Acta Montanistica Slovaca journal, 8(4).

Sun, D., Hoo, K. A. (1999). Dynamic transition control structure for a class of SISO nonlinear systems. IEEE Transactions on Control Systems Technology, 7(7), 622–629.

Theilliol, D., Noura, H. and Ponsart, J. C. (2002). Fault diagnosis and accommodation of a three-tank-system based on analytical redundancy. ISA Transactions, 41(3), 365–382.

Thumati, B. T., Jagannathan, S. (2010). A model-based fault-detection and prediction scheme for nonlinear multivariable discrete-time systems with asymptotic stability guarantees. IEEE Transactions on Neural Networks, 21(3), 404-423.

Tudoroiu, N., Khorasani, K. (2007). Satellite fault diagnosis using a bank of interacting kalman filters. IEEE Transactions on Aerospace and Electronic Systems, 43(4), 1334-1350.

Tudoroiu, N., Sobhani-Tehrani, E., Khorasani, K. (2006). Interactive bank of unscented kalman filters for fault detection and isolation in reaction wheel actuators of satellite attitude control system. Paper presented at the IECON Proceedings (Industrial Electronics Conference), 264-269.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S. N. (2003a). A review of process fault detection and diagnosis part I: Quantitative model-based methods. Computers and Chemical Engineering, 27(3), 293-311.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N. (2003b). A review of process fault detection and diagnosis part II: Qualitative models and search strategies. Computers and Chemical Engineering, 27(3), 313-326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., Yin, K. (2003c). A review of process fault detection and diagnosis part III: Process history based methods. Computers and Chemical Engineering, 27(3), 327-346.

Wang, X., Syrmos, V. L. (2008). Fault detection, identification and estimation in the electro-hydraulic actuator system using EKF-based multiple-model estimation. Paper presented at the 2008 Mediterranean Conference on Control and Automation-Conference Proceedings, MED'08, 1693-1698.

Wei, X., Lui, H., Qin, Y. (2011). Fault isolation of rail vehicle suspension systems by using similarity measure, service operations, logistics, and informatics (SOLI). 2011 IEEE International Conference, 391-396.

Wei, X., Jia, L., Liu, H. (2012). Data-driven fault detection of vertical rail vehicle suspension systems. UKACC International Conference, 589-594.

Willsky, A. S. (1976). A Survey of design methods for failure detection in dynamic systems. Automatica, 12(6), 601–611.

Willsky, A. S., Jones, H. L. (1974). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. IEEE Trans. Automatic Control, 21, 108-112.

Wise, B. M., Gallagher, N. B. (1996). The process chemometrics approach to process monitoring and fault detection. J. of Process Contr. 6(6), 329-348.

Witczak, M. (2007). Modelling and estimation strategies for fault diagnosis of non-Linear systems, from analytical to soft computing approaches. Lecture notes in control and information sciences. Springer, Berlin.

Wu, H., Huang, D. (2008). Improved kernel PCA based on wavelet for fault detection. Paper presented at the Proceedings - ISECS International Colloquium on Computing, Communication, Control, and Management, CCCM, 2, 8-12.

Xia, S. X., Niu, Q., Zhou, Y., Zhang, L. (2008). Mine-hoist fault-condition detection based on the wavelet packet transform and kernel PCA. Journal of China University of Mining and Technology, 18(4), 567-570.

Yang, Q. (2004). Model-based and data driven fault diagnosis with applications to process monitoring. Ph.D. Thesis, Department of Electrical Engineering and Computer Sciences, Case Western Reserve University, USA.

Yang, N., Tian, W., Jin, Z. (2006). An interacting multiple model particle filter for manoeuvring target location, Measurement Science and Technology, 17 (6), 1307-1311.

Yen, G. G., Ho, L. (2003). Online multiple-model-based fault diagnosis and accommodation. IEEE Transactions on Industrial Electronics, 50(2), 296-312.

Yin, L., Zhang, M. (2012) Minimum entropy fault detection for dynamic multivariate nonlinear non-Gaussian stochastic systems. Control and Decision Conference (CCDC), 24th Chinese, 4041 (4046), 23-25.

Zajac, M. (2011). Adaptive particle filter for fault detection and isolation of mobile robots. Towards Autonomous Robotic Systems, 376-377

Zhang, H., Tangirala, A. K., Shah, S. L. (1999). Dynamic process monitoring using multiscale PCA. In “Proc. Of 1999 Canadian Conf. On Electrical and Comput. Eng., 1579-1584.

Zhang, Y., Li, X. R. (1998a). Detection and diagnosis of sensor and actuator failures using IMM estimator. IEEE Transaction On Aerospace and Electronic Systems, 34(4), 1293-1311.

Zhao, G., Jiang, D., Jinhui Diao, J. (2004). Application of wavelet time-frequency analysis on fault diagnosis for steam turbine. 5th International Conference of Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques, CETIM, Senlis, France.

Zhou, J. H., Zhang, D. H., Ooi1, Luo, C. H., Mao, M. S., Wang, D. W. (2010). Comparative study of data driven and model based approaches of rotary machines fault detection and diagnosis. SIMTech technical reports, 11(4), 195-201

Zolghadri, A., Henry, D., Monsion, M. (1996). Design of nonlinear observers for fault diagnosis: A case study. Control Engineering Practice, 4(11), 1535-1544.